62nd ISI World Statistics Congress

ISI 2019 Kuala Lumpur

18 - 23 August 2019

PROCEEDING

INVITED PAPER SESSION

VOLUME 2

62nd ISI WORLD STATISTICS CONGRESS 2019

18 - 23 August 2019, Kuala Lumpur

Come | Connect | Create

www.isi2019.org

PROCEEDING

ISI WORLD STATISTICS CONGRESS 2019

INVITED PAPER SESSION (VOLUME 2)

Published by:
Department of Statistics Malaysia
Block C6, Complex C
Federal Government Administrative Centre
62514 Putrajaya
MALAYSIA

Central Bank of Malaysia
Jalan Dato' Onn
P.O. Box 10922
50929 Kuala Lumpur
MALAYSIA

Malaysia Institute of Statistics
Department of Mathematical Sciences
Faculty of Sciences and Technology
43600 UKM Bangi, Selangor
MALAYSIA

Portal : https://www.isi2019.org
Email : lpc@isi2019.org

Published in February 2020

Suggested citation:
Department of Statistics Malaysia (DOSM). 2019. Proceeding of the 62nd ISI World Statistics Congress 2019: Invited Paper Session: Volume 2, 2019. 495 pages

Preface

The 62nd International Statistical Institute World Statistics Congress (ISI WSC 2019) has a long tradition since 1887, held for the first time in Kuala Lumpur, Malaysia on 18 to 23 August 2019. ISI WSC 2019 is a global gathering of statistical practitioners, professionals and experts from industries, academia and official authorities to share insights in the development of statistical sciences.

The congress attracted an overwhelming number of participants across the regions. The scientific sessions were delivered over five days with parallel sessions and e-poster sessions running all day long. The scientific program reaches across the breadth of our discipline that comprised of Invited Paper Sessions (IPS), Special Topic Sessions (STS) and Contributed Paper Sessions (CPS). Papers presented exhibit the vitality of statistics and data science in all its manifestations.

I am very honoured to present the proceedings of ISI WSC 2019 to the authors and delegates of the congress. The proceedings contain papers presented in IPS, STS and CPS which were published in fourteen (14) volumes. Scientific papers were received from August 2018 and were carefully reviewed over few months by an external reviewer headed by Scientific Programme Committee (SPC) and Local Programme Committee (LPC). I am pleased that the papers received cover variety of topics and disciplines from across the world, representing both developed and developing nations.

My utmost gratitude and appreciation with the expertise and dedication of all the reviewers, SPC and LPC members for their contributions that helped to make the scientific programme as outstanding as it has been.

Finally, I wish to acknowledge and extend my sincere thanks to the member of National Organising Committee of ISI WSC 2019 from Department of Statistics Malaysia, Bank Negara Malaysia, Malaysia Institute of Statistics and International Statistical Institute for endless support, commitment and passion in making the ISI WSC 2019 a great success and congress to be remembered.

I hope the proceedings will furnish the statistical science community around the world with an outstanding reference guide.

Thank you.

Dr. Mohd Uzir Mahidin
Chairman
National Organising Committee
62nd ISI WSC 2019

Scientific Programme Committee of the 62nd ISI WSC 2019

Chair
Yves-Laurent Grize, Switzerland

Vice-Chairs
Kwok Tsui, Hong Kong, China
Jessica Utts, USA

Local Programme Committee Chair and Co-Chair
Rozita Talha, Malaysia
Ibrahim Mohamed, Malaysia

Representatives of Associations
Rolando Ocampo, Mexico (IAOS)
Cynthia Clark , USA (IASS)
Bruno de Sousa, Portugal (IASE)
Sugnet Lubbe, South Africa (IASC)
Mark Podolskij, Denmark (BS)
Beatriz Etchegaray Garcia, USA (ISBIS)
Giovanna Jona Lasinio, Italy (TIES)

At-Large Members
Alexandra Schmidt, Canada/Brazil
Tamanna Howlader, Bangladesh

Institutional/Ex Officio
Helen MacGillivray, ISI President, Australia
Fabrizio Ruggeri, Past SPC Chair, Italy

Liaison at ISI President's Office
Ada van Krimpen, ISI Director
Shabani Mehta, ISI Associate Director

Local Programme Committee of the 62nd ISI WSC 2019

Chairperson
Rozita Talha

Co-Chairperson
Prof. Dr. Ibrahim Mohamed

Vice Chairperson
Siti Haslinda Mohd Din
Daniel Chin Shen Li

Members
Mohd Ridauddin Masud
Dr. Azrie Tamjis
Dr. Chuah Kue Peng
Dr. Tng Boon Hwa
Prof. Dr. Ghapor Hussin
Prof. Madya Dr. Yong Zulina Zubairi
Prof. Madya Dr. Wan Zawiah Wan Zin @ Wan Ibrahim
Dr. Nurulkamal Masseran
Siti Salwani Ismail
Rosnah Muhamad Ali
Dr. Ng Kok Haur
Dr. Rossita Mohamad Yunus
Muhammad Rizal Mahmod

TABLE OF CONTENTS

Invited Paper Session (IPS): Volume 2

Communicating on Lebanese inflation: Deciphering the impact of imported inflation

Sana Antoine Souaid Jad
Banque du Liban, Beirut, Lebanon

Abstract

The main purpose of this paper is to estimate the importance of imported inflation in the total headline inflation computed using the Consumer Price Index for the Lebanese economy. For decision-makers, it is crucial to disentangle the effect of foreign exogenous factors from domestic ones on headline inflation in order to conduct the monetary policy optimally. Since Lebanon's consumption is mainly based on imports and highly exposed to international price fluctuations, imported inflation plays a prominent role and its amplitude should be measured.

Two different approaches have been followed to estimate the share of imported inflation: a sectorial approach and a model estimation. In the first approach, we use the structure of the Lebanese Consumer Price Index (contribution of each component in the final CPI basket) in order to evaluate the share of variation due to the externally driven items. In this context and given the big share of imported volatile items such as food and energy and their significant contribution in the inflation process, a closer analysis between headline and core inflation is performed.

In the second approach, we estimate an econometric model[1] where changes in CPI originate from domestic factors (money supply, interest rates) and external ones (Brent price, effective exchange rate[2]). The share of inflation volatility due to external factors is computed and compared to the results found in the first approach.

As expected, both approaches show that imported inflation plays a major role in headline inflation since it accounts for 50% of the total volatility of headline inflation as per approach 1 and around 40% based on approach 2. These results prove that Lebanon is quite vulnerable to external shocks which can have significant repercussions on the monetary policy of the country.

Keywords

headline inflation; core inflation; imports; exchange rates; Brent prices.

---

[1] Based on the BDL Statistics and Economic Research Department's current inflation model.

[2] The Lebanese Pound is pegged to the US Dollar meaning that fluctuations in the EER are essentially due to variations in the USD/EUR exchange rate (EU is the main trading partner of Lebanon). Thus, a shock on EER can be considered completely exogenous to the Lebanese economy.

1. Introduction

The relationship between imports and inflation has been the subject of many studies. It is clear that a major part of inflation fluctuations arises from external factors which cannot be controlled by the local monetary authority. Consequently, policymakers need to assess the true impact of their actions on the inflation level and estimate the volatility caused by external shocks. This has led monetary authorities and academics to define core inflation as a new measure of inflation in which the influence of external and highly volatile factors has been reduced (ideally removed). Such instrument enables policymakers to isolate and target the structural and controllable[3] factors behind inflation. In other words, it captures the component of price change that is common to all domestic items and excludes components that are subject to frequent and temporary price shocks (such as food and energy prices).

This paper aims to analyze both headline and core inflation in the Lebanese economy, and helps identify to what extent the Lebanese headline inflation is affected by domestic factors and imported inflation.

The next section describes the methodology which has been followed to measure imported inflation: two stand-alone approaches, namely a sectorial empirical approach and an econometrical one, are used to estimate the share of imported inflation in headline inflation. By adopting these two distinct approaches, we are able to perform a robustness check to the overall study. The following section presents the results of our approaches while the last section concludes the study and provides us some final remarks.

2. Methodology

a- Sectoral approach

The first approach is based on the way the Consumer Price Index (CPI) is built: a choice of items which are highlydependent on external factors is performed and allows to assess their impact on CPI and to build a core CPI measure (by removing these items). For the sake of completeness, a brief review of how the CPI is computed can be found hereafter.

The CPI is an index of the cost, through time, of a fixed market basket of goods and services purchased by a typical household for consumption in some base period. The composition of the Lebanese basket is derived from a detailed expenditure survey conducted by the Central Administration of

[3] Through expansionary or contractionary monetary policies.

Statistics at the level of households and individuals. The CPI is computed by applying the Laspeyres index which uses historical quantities (weights):

$$P_t = \frac{\sum_{i=1}^{n} q_{i0} p_{it}}{\sum_{i=1}^{n} q_{i0} p_{i0}} \times 100$$

q_{i0} is the base period quantity of good i
p_{i0} is the base period price of good i
p_{it} is the time t price of good i

Where each subindex i is related to a given item of the basket. The choice of weights q_i is fixed and changes to CPI occur through changes in the price levels p_i. Headline inflation refers to the rate of change in the CPI. It captures the changes in the cost of living based on prices fluctuations of items in the basket of commodities and services consumed by households. However, using the headline inflation rate for monetary policy analysis is sometimes misleading due to the volatility of certain components included in the CPI. The current composition of the CPI basket and the weight for each one of the items can be found in the following figure.

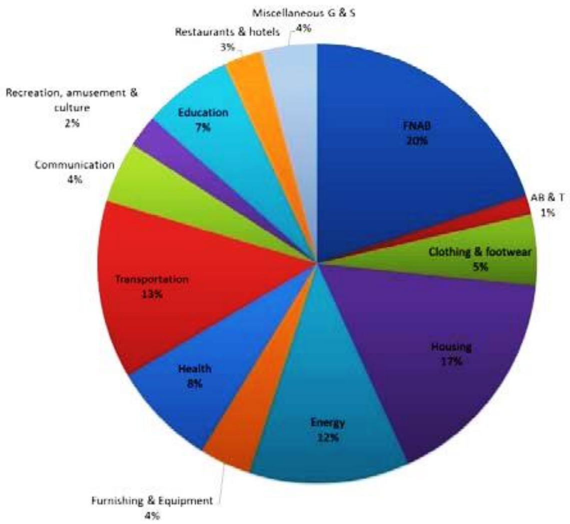


Figure 1: composition of CPI Sources: CAS

Based on the previous formula, we can compute the volatility of CPI due to each of sectorial items. Two factors come into play in determining the impact of each item:

- **Weight**: the higher the weight of the item in the CPI formula, the more impact it will have on headline CPI
- **Volatility of prices per sector**: the higher the variations of prices in a given sector (compared to other sectors), the higher the influence of this sector on headline CPI

The influence of each component on the headline inflation is simply the product of its weight by its year-on-year change. The link between imported inflation and this approach comes from the fact that prices in some of the sectors are exclusively determined by the outside world, such sectors include energy and food items. Consequently, a first assessment of the impact of imported inflation can be made by computing the share of volatility in CPI due to these components.

Furthermore, an alternative for constructing a core CPI measure is to exclude the "external" items from headline CPI and re-weight the remaining items on a prorata basis (exclusion method).The new core price level includes goods 1 to m, and fully excludes goods m+1 to n. The denominator rescales the weights for commodities 1 to m and commodities m+1 to n are re-assigned to have weights zero.

$$\mathrm{P}_{\mathrm{t}}^{\mathrm{core}} = \frac{\sum_{i=1}^{m} w_i \frac{P_{it}}{P_{i0}} \times 100}{\sum_{i=1}^{m} w_i}$$

Where m≤n

Several other methods are used to compute core inflation: Trimmed mean, principal component analysis, volatility weights, and permanent and variable exclusion approaches[4].The exclusion method is the most common approach (used by several Central Banks) due to its simplicity. For the case of Lebanon, a core CPI excluding Energy and Food has been computed and applied to the CPI series calculated by the Central Administration of Statistics (CAS) since 2008. It excludes both Food and Energy categories (with a weight of 31.8%) from the CPI index.

[4] A detailed presentation of these methods can be found in an IMF Staff paper written by Mick Silver (check references).

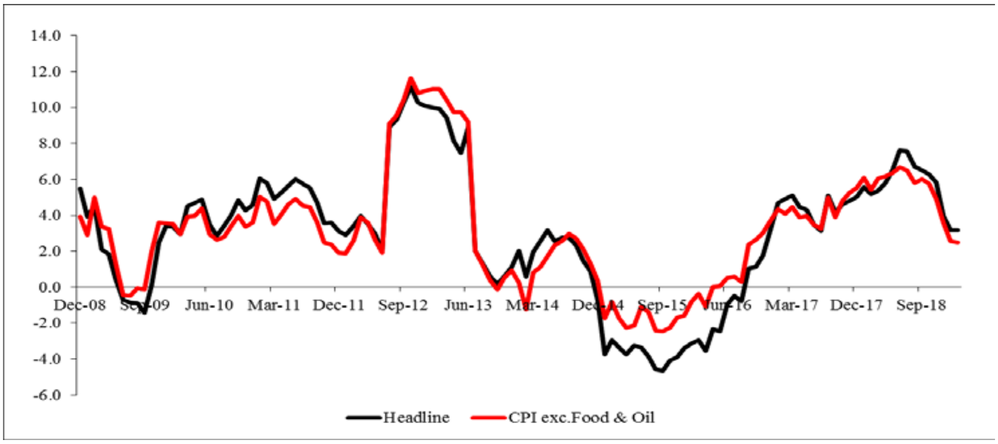


Figure 2: Headline versus Core CPI (year on year change) Sources: CAS, BDL

Based on figure 2, we can notice that headline CPI is more influenced by oil prices and, consequently, more volatile than core CPI. For instance, the jump in prices which happened in 2009 is essentially due to the Brent price increase. As expected, the core CPI is slightly less impacted by the shock on oil prices and stands below headline CPI during the period 2009-2011. The opposite happened in 2014, when Brent prices dropped drastically: the drop of core CPI is smaller than the one of headline CPI.

Nevertheless, we also notice that core CPI is only a partial improvement to headline CPI since the effects of oil prices are still clearly seen in the graph. Actually, this approach relies on the fact that components are classified into two clear separate groups: those impacted by domestic monetary measures and those exclusively influenced by external factors. Unfortunately, this assumption does not hold for many reasons:

- Some items are impacted by external and internal factors. For instance, transportation prices are clearly impacted by the world price of oil as well as by the level of wages in Lebanon (bus driver, mechanic).
- Cost of imports can be transmitted to all items through indirect or second-round effects. For instance, an increase in the price of imported capital machinery and raw materials affect indirectly all industry-related items of CPI.

b- Model-based approach

In order to overcome the previous limitations, a regression model is proposed to estimate the importance of imported inflation in the total headline inflation in Lebanon. This model does not make any distinction between sectors but explains the variations in CPI using variables which describe both domestic and external economic environment.

The main external variables impacting consumer prices are:

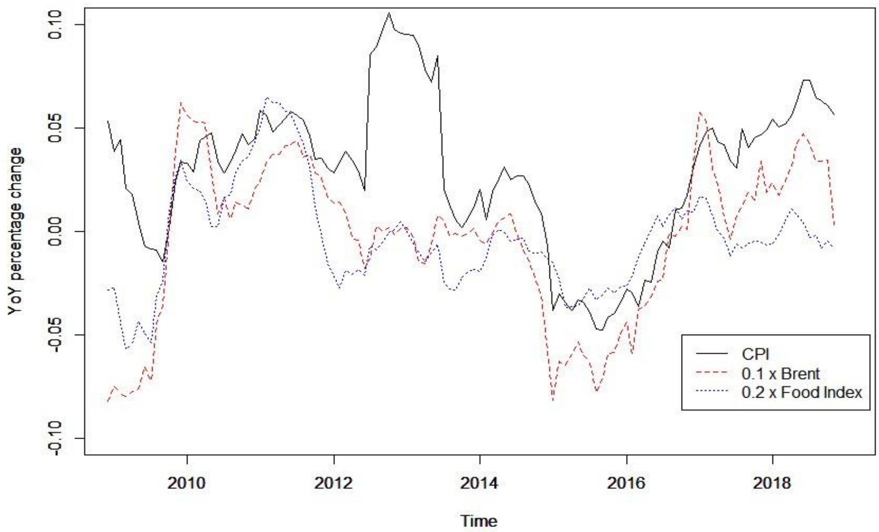
- Brent price (US Energy Information Administration)
- Agricultural Index (World Bank)
- USD/EUR exchange rate (FED)

The domestic variables are the following:

- M3: money supply (BDL)
- WAIR: weighted average interest rate (BDL)

A dummy has been included for the year 2012 to account for the "jump" in CPI due to the change in its housing component. In addition to this "artificial" change, the year 2012 witnessed an increase of salaries in the public sector as well as a massive arrival of Syrian refugees which translated into higher demand and prices in the housing sector. As we can notice from figure 3, the variations in CPI closely follow the variations of the Brent price and Food Index.

Figure 3: CPI, Brent and Food index Sources: CAS, US EIA, WB Commodity



Price Data

OLS estimation

Based on unit root tests, we decide to consider the first log-differenced series for the OLS estimation. Furthermore, due to the highly noisy nature of the month-on-month log-differenced series, we perform our estimation using year-on-year logdifferences. Finally, the agricultural index and the Brent price are strongly correlated (0.75) which lead us to keep the Brent time series for the final estimation.

The following OLS regression has been performed:

$$\Delta(\log(CPI_{headline})) = \beta 0 + \beta 1\ \Delta(\log(M3)) + \beta 2\Delta(\log(WAIR)) + \beta 3\Delta(\log(BRENT)) + \beta 4\Delta(\log(EXR)) + dummy + \epsilon$$

In addition to the headline CPI, we performed the same regression on a measure of the "core CPI" which excludes Food
and Energy items. Such regression allows us to check that, when considering the core CPI measure, the impact of external variables is reduced.

Estimating each variable's importance

When explanatory variables are uncorrelated, the contribution of each variable in explaining the variations of CPI (improving R2) is unambiguous. However, when these variables become correlated, it is no longer clear how each one is impacting CPI. We decided to adopt the LMG (Lindeman, Merenda, and Gold) measure for assessing the relative importance of each variable in the linear regression model. The way this measure works is pretty simple: adding a variable (WAIR, for example) to the regression improves the goodness-of-fit R2 metric by a given amount (R2 jumps from 0.49 to 0.55, improvement = 0.06). When the explanatory variables are correlated, the amount of such an improvement (0.06) depends on which variables are already included in the model before we added the new variable (WAIR). The LMG measure averages these improvements over all possible orderings of regressors. In general, the LMG measures do not add up to 1 since variations in CPI are not perfectly explained by our variables.

3. Results

The results of the sectorial approach are presented in figure 4 and it can be pointed out that:

- The portion of Lebanese inflation that comes from non-tradable items such as health, housing and education represents only 26.6%. These items are purely domestic and are not directly impacted by global trade. The majority of other items are tradable especially food and energy items.
- Food and non-alcoholic beverages constitute a major component of Lebanon's consumer price index, with the highest weight of 20%. The food price index in Lebanon has been increasing faster than the consumer price index indicating rising food prices. In fact, food prices have increased by an average annual rate of 5% during the period 2010-2013 against 2% during the period 2012-2018. This comes in line with trends in international food prices which also increased by an average annual rate of 6% during the same period. Since Lebanon imports most of its food demand, it is consequently highly exposed to international food price fluctuations.
- The items that contribute the most to inflation are Energy (36.1%), Food and non-alcoholic beverages and Housing (15.9% each). Such a result confirms our main intuition that Food and Energy prices fluctuations have significant repercussions on headline inflation. Even though the share of the energy item in the CPI basket is only equal to 11.8%, it explains 36.1% of the volatility of headline CPI. If we add to that the transportation item which is directly impacted by the price of fuel, the direct effects of oscillations in oil prices explains 49.8% of the

volatility of headline CPI. This result confirms that imported inflation, in particular the one due to variations in world oil prices, is the main driver of headline CPI inflation in Lebanon.

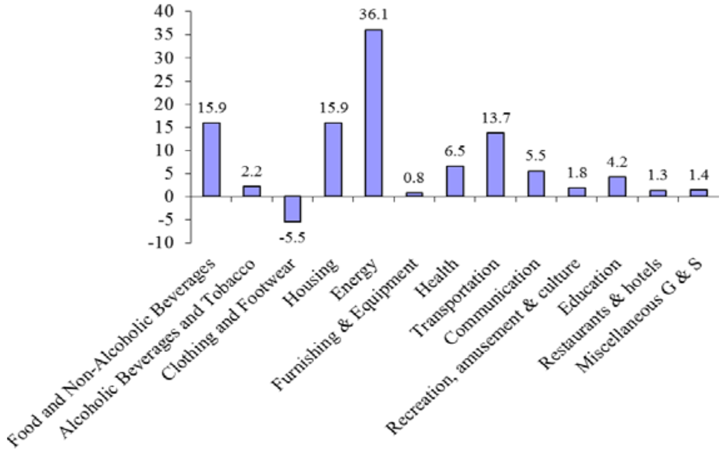


Figure 4: average components contributions to inflation between 2012 and 2018

As opposed to the sectorial analysis, the econometric model quantifies the impact of external factors regardless of the sectors through which these shocks are transmitted. Consequently, the second approach complements and provides a robustness check to our initial results. The estimation of the linear model using the log-differenced headline CPI as the explained variable yields the following result:

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.016152	0.004957	3.258	0.00148 **
diff(log(m3), lag = 12)	0.104585	0.054642	1.914	0.05813 .
diff(log(wair), lag = 12)	0.074915	0.030839	2.429	0.01669 *
diff(log(brent), lag = 12)	0.045214	0.005968	7.576	1.02e-11 ***
diff(log(exr), lag = 12)	-0.027909	0.025065	-1.113	0.26784
dummy	0.069336	0.005625	12.326	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.01825 on 114 degrees of freedom
Multiple R-squared: 0.6828, Adjusted R-squared: 0.6689
F-statistic: 49.08 on 5 and 114 DF, p-value: < 2.2e-16

The Brent price is clearly highly significant and confirms that the world oil prices are a major driver of Lebanese inflation. The model estimates that a 1% increase in the price of oil leads to an increase of 0.065% in inflation, a very strong result given the highly volatile nature of the Brent price level.

In addition to the headline CPI, we performed the same regression on a measure of the "core CPI" which excludes Food and Energy items. Such regression allows us to check that, when considering the core CPI measure, the impact of external variables is reduced.

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.016152	0.004957	3.258	0.00148 **
diff(log(m3), lag = 12)	0.104585	0.054642	1.914	0.05813 .
diff(log(wair), lag = 12)	0.074915	0.030839	2.429	0.01669 *
diff(log(brent), lag = 12)	0.045214	0.005968	7.576	1.02e-11 ***
diff(log(exr), lag = 12)	-0.027909	0.025065	-1.113	0.26784
dummy	0.069336	0.005625	12.326	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01825 on 114 degrees of freedom

Multiple R-squared: 0.6828, Adjusted R-squared: 0.6689

F-statistic: 49.08 on 5 and 114 DF, p-value: < 2.2e-16

As expected, the elasticity of Brent decreases to 0.045 but this variable remains very significant proving that the exclusion of the Food and Energy items is not enough to eliminate the indirect effects of oil prices. Furthermore, we notice that the coefficients for interest rates and money supply become significant at the 5% and 10% levels respectively. In addition to that, their elasticities are higher than the ones found for headline inflation meaning that core inflation reflects more distinctly the impact of monetary policy than headline inflation.

Measure	Headline	Core 3 (excl. food & energy)
Δ(log(M3))	0.3 %	0.4 %
Δ (log(WAIR))	0.1 %	0.9 %
Dummy	26.6 %	41.8 %
Δ(log(Brent))	31.5 %	20.3 %
Δ(log(EXR))	7.2%	4.9 %

Table 1: LMG measure of relative importance

Finally, the estimation of the LMG measures can be found in table 2, and the following remarks could be made:

- The Brent price explains more than 30% of the variation in the headline CPI. When considering core CPI which excludes the energy item, this participation is significantly reduced. Nevertheless, it remains significant which means that Brent price has spillover/indirect effects on other items of CPI (transportation, for example).
- The housing “shock” (modeled by the dummy variable) has a tremendous impact on CPI. This impact increases as we remove the Energy and Food items from CPI.
- WAIR and M3 have a relatively minor impact on CPI. It should be noted that we are considering a model with differenced series. In a model with levels (such a VECM), M3 would be a main driver of CPI (both time series are cointegrated).

4. Conclusion

Both approaches lead to the same clear conclusion: Lebanese inflation is mainly driven by external factors, in particular the world price of oil. Such external shocks affect directly and indirectly most items in the CPI consumption basket. Despite a small decrease in its coefficient compared to headline CPI, the influence of the Brent price on core CPI remains significant. As expected, the impact of monetary policy on inflation is more significant when considering core inflation. This study confirms that the small, open and oil-importing economy which is Lebanon is extremely sensitive to external factors such as world oil prices when it comes to inflation. Such external factors account for approximately half of the volatility of inflation. Despite its limitations, the core inflation measure is a better alternative than headline inflation when it comes to conducting monetary policy since it filters out the direct effects of external shocks and reflects more accurately the impact of money supply on inflation.

References

1. Corrigan, T. D. (2005). The Relationship Between Inport Prices and Inflation in the United States. WCOB Faculty Publications, Paper 18. Retrieved from http://digitalcommons.sacredheart.edu/wcob_fac/18
2. Hamilton, J. D. (2012, March). Import Prices and Inflation. International Journal of Central Banking, 8(1), 271-279.
3. Jad, S. S. (2013). Measuring Core Inflation for Lebanon, www.bdl.gov.lb.
4. Johnson, M. (n.d.). Core Inflation: A measure of inflation for policy purposes. Montreal.

5. McCarthy, J. (1999). Pass-Through of Exchange Rates and Import Prices to Domestic Inflation in some Industrialized Economies. New York: Federal Reserve Bank of New York.
6. Mishkin, G. F. (2007, October 20). Headline versus Core Inflation in the Conduct of monetary Policy. Montreal, Canada: The Federal Reserve Board.
7. Motaza, M. G., & Farah, N. (2009). Estimating Inflation Rates of import-Concentrated Commodities.
8. Muktadir-Al-Mukit, D., Shafiullah, A. Z., & Ahmed, M. (2013, 1 1). Inflation Led Import or Import Led Inflation: Evidence from Bangladesh. Asian Business Review, 2(4), 2/2013.
9. Silver, M. (2007). Core inflation: Measurement and statistical issues in choosing among alternative measures. IMF Staff papers 54.1 : 163-190.

The importance of communication between compilers and users in improving the quality of Statistics: The bank of Japan's experience

Kuniko Moriya
Bank of Japan, Tokyo, Japan

Abstract

Economic and financial statistics form the basis of assessments and forecasts of economic and financial conditions. In order for central banks and other policymakers to conduct policies appropriately, ongoing improvement in the quality of these statistics is essential. In relation to these statistics, the Bank of Japan wears two hats — statistics compiler and user.

As a compiler of statistics, the Bank makes continuous efforts to improve the quality of the statistics it compiles, reflecting changes in economic and financial structures in a timely manner. It explores better data sources, develops more accurate and effective estimation methods, and expands the coverage of its statistics where necessary and possible. When the Bank decides to make major revisions of statistics, it provides users of these statistics with the necessary information regarding the revision plans in advance and collects opinions and comments through a public consultation process. Taking on this feedback, the Bank aims to provide the most useful statistics possible for users. Central banks also use a wide range of statistics for the assessment of economic and financial conditions. The quality of these statistics is important to make assessments and policy decisions. In Japan, statistics reform has been in progress under the Japanese government's Statistics Reform Promotion Council, with the aim of promoting evidence-based policy making (EBPM). In this context, the Bank of Japan contributes to improvements in statistics not only as a compiler but also as a user of them.

This paper outlines the Bank's efforts to improve the quality of statistics, with an emphasis on the importance of communication between compilers and users. The paper first provides an overview of the Bank's communication with users of the statistics it compiles. Examples are presented, such as those reflecting user feedback on comprehensive revisions or updates of its statistics. Secondly, the Bank's experience of improvements in statistics which are brought about by the synergy between statistics compilers and users are presented. The paper also explains the Bank's contributions to statistics reform, which Japan's current administration promotes.

Keywords

official statistics; public consultation process; statistics reform

1. Introduction

Economic and financial statistics form the basis of assessments and forecasts of economic and financial conditions. Ongoing improvement in the quality of these statistics is essential for policymakers, including central banks, to conduct their policies appropriately. For these statistics, the Bank of Japan takes on two roles, each with a different perspective — as a compiler and user of statistics. As a compiler of statistics such as the producer price indexes for goods and services, the flow of funds account, and the Tankan (Short-Term Economic Survey of Enterprises in Japan), the Bank is continuously working to improve the quality of its statistics. At the same time, the Bank uses a wide range of statistics in making policy decisions. This paper is organized as follows. Section 2 presents the Bank's activities as a compiler of statistics. The Bank communicates with users of the statistics it compiles. On the process of recent substantial revisions of the Services Producer Price Index (SPPI) and the Tankan, communications with users will be explained in detail as examples. Then, Section 3 explains the Bank's experience of improvements in statistics which are brought about by a synergy between statistics compilers and users. The Bank's contributions to the government's statistics reform are also presented. Section 4 concludes the paper.

2. The Bank's communications with users of its statistics

The Bank of Japan (BOJ) compiles a variety of statistics, such as financial statistics, price indexes, and the Tankan, which indicates corporate sentiment. When the Bank makes substantial revisions to its statistics, it announces the changes in advance. By doing so, the Bank ensures a broad range of users and reporting entities can provide comments and opinions, in line with the third item of "The Basic Principles for the Compilation, Release, and Development of Statistics" (hereafter referred to as the Basic Principles) published in 2009.

In the following sections, we introduce the Basic Principles and their background, and then the communications with users which took place when substantial revisions were being made to the Services Producer Price Index (SPPI) and the Tankan are presented as examples.

2-1. Basic Principles for statistics at the Bank of Japan

The Basic Principles outline the Bank's basic principles for the compilation of statistics and form a quality assurance system. The principles cover various aspects of statistics management across seven categories, as in Table 1. The motivation to compile the Basic Principles was from the first comprehensive revision of the Statistics Act (Act No.18 of 1947) in 60 years, which promoted an improvement in convenience for users and in transparency.

[Table 1] Basic Principles for statistics at the Bank of Japan

1. Provision of accurate and appropriate statistics
2. Improvement of convenience for user of statistics
3. Enhancement of transparency
4. Reduction of the burden on reporting firms and provision of detailed explanations to obtain their cooperation with the survey
5. Maintenance of strict confidentiality in data management
6. Rationalization of and improvement in the efficiency of operations for compiling statistics
7. Mutual cooperation and appropriate sharing of roles with relevant organizations such as administrative bodies

Three (No. 1, 2, and 3) of the above seven Basic Principles refer to the improvement of statistics quality from the various user perspectives. The third principle (No. 3) is the enhancement of transparency, comprising three points: (1) making public the rules pertaining to the methods of release; (2) promptly releasing any corrections to statistics; and (3) listening to a variety of external opinions regarding the substantial revision of statistics. As in (3), when making substantial revisions to BOJ statistics, the Bank announces the revision contents in advance in order to receive opinions and comments from a broad range of users and reporting entities. In practice, however, it is often difficult for the Bank to incorporate all the opinions and requests it receives, for various reasons. Even in such cases, the Bank discloses these opinions and requests, as well as its reaction to them, in order to increase the transparency of the consultation process.

2-2. Recent improvements of statistics compiled by the Bank of Japan

In recent times, the Bank has made improvements to a variety of statistics, and continues to work on them, as shown in Table 2.

[Table 2] Recent improvements of BOJ Statistics

Monthly/year	Statistics	Revisions made
Mar, 2014	Balance of Payments (BOP)	Adoption of BPM6
Jun, 2014	Services Producer Price Index (SPPI)	Base year revision (from 2005 to 2010)
Mar, 2016	Flow of Funds Accounts (FFA)	Adoption of 2008SNA
Feb, 2017	Corporate Goods Price Index (CGPI)	Base year revision (from 2010 to 2015)
Mar, 2014, Mar, 2015, Mar, 2017, Mar, 2018, and around 2020	*Tankan* (Short-Term Economic Survey of Enterprises in Japan)	Periodic revision of sample enterprises and examination of sampling errors. Revisions of survey items, sample design method, etc. See further detail in 2-2-2.
Jun, 2019	Services Producer Price Index (SPPI)	Base year revision (from 2010 to 2015)

2-2-1. Services Producer Price Index (SPPI)

The Services Producer Price Index (SPPI) is a price index, based on a statistical survey which measures movements in the prices of service products traded in the corporate sector. The survey is conducted by the Bank in accordance with the Statistics Act. The purpose of the index is to facilitate a better understanding of supply and demand conditions for individual service products, as well as to provide materials for assessing the state of the economy and making decisions on monetary policy. The SPPI is also used as a deflator in calculating real values — i.e., in removing price fluctuations from nominal value — such as in the estimate of the System of National Accounts (SNA, hereafter referred to as GDP statistics) compiled by the Cabinet Office, and cited as a reference for price-setting activities.

The calendar year 2015 is used as the base year for both the index and weight calculation of the current SPPI. The base year is updated every five years to ensure the SPPI remains appropriate and accurate, and to improve its quality as much as possible. The Bank updated its base year from 2010 to 2015 in June 2019. In the rebasing process, the Bank first published the Basic Policy (available in Japanese only) in May 2018, and asked for feedback. After taking into account the wide range of opinions and proposals received, the Bank published the Final Draft (available in Japanese only) in March 2019. The Bank listed three points as priority themes for the SPPI revision to the 2015 base: (1) response to changes in economic and industrial structures; (2) initiatives to improve the accuracy of indexes; and (3) response to the government's statistics reform and the needs of deflators.

Following the experimental survey, the Bank has begun a full-scale survey of the “Wholesale Services Price Index (WSPI),” from the 2015 base SPPI. This also responds to the increasing need for more accurate GDP statistics in Japan through the improvement of the service sector’s statistical development.

The wholesale trade sector is one of the most significant sectors in the Japanese economy. Wholesale products come to about 437 trillion yen (2016, Economic Census), and the nominal value added to the whole economy is 8.4% (2015, GDP statistics). Thus wholesale services should be in the scope of the SPPI. However, due to the difficulty in surveying wholesale service prices, up until the most recent SPPI revision this time, the Bank has regarded them as “uncovered services” in the SPPI, and incorporating them into the SPPI has been an outstanding issue. In addition to measuring the price trends of wholesale services, there are several other benefits which the WSPI will yield, such as: (a) accurate deflators in GDP statistics, (b) enabling the analysis of productivity of the wholesale trade sector, (c) enabling assessment of economic conditions and analysis of the price setting behavior of wholesalers. It is because of these benefits that economists and external statistics authorities such as the Cabinet Office have called for the creation of the WSPI for many years.

The WSPI has just been adapted to the GDP statistics as its deflators from the quarterly estimates (first preliminary estimates) for April-June 2019, published in August 2019. In the GDP statistics, the nominal output of the wholesale trade sector is measured by the nominal amount of wholesale margin (the difference between value of total sales and that of total purchase), and in order to calculate real output of wholesale trade sector, it is necessary to divide the nominal margin by a “wholesale deflator.” Since the wholesale trade sector consists of large amount of nominal GDP as mentioned above, an accurate deflator for the wholesale trade sector would be indispensable for the precise measurement of real GDP. Previously, in Japan’s GDP statistics, the wholesale deflator was simply estimated using price data of products purchased by wholesalers (i.e. the Producers Price Index) as a substitute for the corresponding service prices. This method, is however, often mentioned as a less appropriate method because of the strong assumption that price trends of wholesale services are identical to price trends of corresponding goods transacted by wholesalers. In contrast, in the U.S. and Canada, the price statistics divisions compile quality-adjusted price indexes of wholesale services by surveying prices with constant quality from a large number of companies, and the indexes are used to calculate wholesale deflator in their GDP statistics.

In other countries such as the United Kingdom and Germany, where WSPI does not exist, real output of the wholesale trade sector is estimated using the assumption that the volume of margins follows the volume of sales (measured by deflating nominal sales using the sales price index). As this assumption is relatively plausible, and it does not require the additional burden of constructing new statistics, many countries have adopted this methodology. However, this method does not take quality of wholesale services into account (United Nations [2010]). The Bank intends to construct the WSPI as a quality-adjusted price index of wholesale services by surveying prices from companies, as is done in the U.S. and Canada. This has satisfied a call of "Basic Policy for the Fundamental Reform of Economic Statistics" (hereafter referred to as Statistics Reform) of December 2016 from the Council on Economic and Fiscal Policy, which suggested that the Bank should start surveying the prices of wholesale services so as to improve the accuracy of the GDP statistics.

2-2-2. The *Tankan*

The Tankan is a statistical survey conducted by the Bank in accordance with the Statistics Act. The Tankan aims to provide an accurate picture of business trends for enterprises in Japan, thereby contributing to the appropriate implementation of monetary policy. The survey is conducted quarterly, in March, June, September, and December. In principle, survey results are released at the beginning of April, July, October, and mid-December.

In order to capture the actual economic conditions accurately and respond to rapidly changing financial and economic structures, the Bank takes great care to ensure the appropriateness of Tankan survey items. In addition, the Bank makes efforts on an ongoing basis to enhance the methodologies used in the compilation of the Tankan. Non-sampling errors are also checked through a number of measures.

In terms of conducting the Tankan survey, the Bank aims to improve user convenience as well as to reduce the burden on survey respondents to the greatest extent possible. While the Bank has added some necessary survey items, it has also removed items that draw relatively less attention from users and overlap with other sets of statistics, as in Table 3.

[Table 3] Changes to survey items

Month/year	Additional items	Removed items
Mar, 2014	Inflation Outlook (Outlook Prices, General Prices)	Material Costs, Personnel Expenses, Depreciation Expenses, Financial income, Financial Expenses, etc.
Mar, 2017	R&D Investment	-
Around 2020	Exchange Rates (Yen/Euro), Overseas Business Activities (Overseas Sales, Overseas Fixed Investment)	Operating Profits, Fixed Investment, Land Purchasing Expenses* Software Investment*

Note: Asterisk (*) indicates a change of survey term from semi-annual to annual basis to obtain fiscal-year figures.

When the Bank considers changing survey items, it first provides users with the revision plans in advance and collects opinions and comments through a public consultation process. After taking into account these user comments, the Bank modifies its revision plan where necessary to ensure that any revision of the Tankan is appropriate. For example, in order to receive feedback regarding its revision plan, the Bank published the Basic Policy (available only in Japanese) in June 2016, which included the prospective expansion of the survey item "Software, R&D, and fixed investment excluding land purchasing expenses" to cover "R&D investment." Through this public consultation process, however, the Bank recognized that users called for the introduction of an independent item, "R&D investment" and the continued publication of the previous item "Software and fixed investment excluding land purchasing expenses", as it was.

[Table 4] Changes to survey items relating to the introduction of R&D

Previous Item	Originally proposed item change	Revised items
-	Software, R&D and fixed investment excluding land purchasing expenses	Software, R&D and fixed investment excluding land purchasing expenses
Software and fixed investment excluding land purchasing expenses	-	Software and fixed investment excluding land purchasing expenses
-	-	R&D investment

The Bank, therefore, published the final draft of the revision in December 2016, and explained its decisions regarding the survey items: (1) the expansion of the survey item "Software, R&D, and fixed investment excluding land purchasing expenses" as originally proposed; (2) the introduction of a new item, "R&D investment"; and (3) the ongoing publication of the previous item "Software and fixed investment excluding land purchasing expenses."

3. The Bank's contribution to the improvement of statistics from a user perspective

This section explains the Bank's contribution to improvements in statistics from a user perspective. A synergy between statistics compilers and users is touched upon, using the example of the revision of Tankan. Other examples of the synergy between statistics compilers and users at the Research and Statistics Department (RSD) — in developing economic indicators as an analytical tool, which can be used to measure economic activity — are also explained. Further, the Bank's contribution to the government's statistics reform is presented.

3-1. The Bank's contribution to the improvement of statistics from a user perspective and a synergy between statistics compilers and users

The Research and Statistics Department at the Bank has an advantage in giving feedback between statistics compilers and users in that its organizational structure is comprised of research divisions and statistics compilation divisions. Within the department, from a user perspective, the research divisions can voluntarily request the improvement of the quality of official statistics[1] compiled by the statistics divisions, such as financial statistics, price indexes, and the Tankan. The statistics divisions then also take into account feedback from research divisions, in addition to those received through the public consultation process, as mentioned in the section 2. After a request is incorporated into the statistics, economists at the research divisions can make use of the improved statistics, ensuring the accuracy of their analysis in order to make appropriate judgements on economic conditions. This internal feedback loop between the research and statistics divisions can produce synergy effects within the RSD.

For example, in the case of Tankan, a proposal was made from the RSD's research division to introduce survey items for inflation outlook. The RSD's statistics division included the proposal in the basic policy for the revision of Tankan. Through a consultation process, the statistics division then decided on the introduction of new survey items for the inflation outlook from the March 2014 survey. Since the new inflation outlook survey items were incorporated into the Tankan, economists at the Bank have frequently used the survey for their analysis examining firms' inflation expectations. In this way, the synergy effect produced between the analytical and statistics compilation divisions through user feedback for the improvement of statistics is incorporated and this facilitates a new type of analysis.

3-2. Other examples of the synergy between statistics users and compilers — development of economic indicators by the research divisions at the Bank

In assessing economic developments, the Economic Research Division (ERD) regularly compiles several indicators, such as the Consumption Activity Index (CAI), and the Real Exports and Imports (REI). The CAI is a comprehensive index which captures private consumption movements as an analytical tool; the REI is an economic indicator compiled in line with exports and imports in the real GDP to measure the real value of export/import goods. The CAI and

[1] "Official statistics" are defined by the Statistics Act as the statistics compiled by the administrative organs, including the Bank of Japan.

the REI are not official statistics but economic indicators for conjunctural assessment, of which compilation method and data sources are changed flexibly for quality improvements on an ongoing basis.

Development of these indicators provides another example of the synergy effects of internal feedback between the research and statistics divisions at the RSD. Both economic indicators use the Corporate Goods Price Index (CGPI) extensively, which is produced by the Price Statistics Division (PSD) in the RSD, as deflators. In the course of developing these indicators, economists at the ERD asked extensively about details of CGPI, for example, regarding its definition, calculation method, and items surveyed, so that the CGPI could be used properly in the CAI and the REI. Such close communication between statistics compiler and user improved the economic indicators compiled by the research divisions. At the same time, statisticians at the PSD better recognized user needs and detected clues for the improvement in quality of the statistics they compile.

3-3. Contribution to the government's statistics reform

In December 2016, the Council on Economic and Fiscal Policy decided on the Statistics Reform, which was then endorsed in May 2017 by the Statistics Reform Promotion Council, where the Governor of the Bank is a member and contributes to its discussions. The Statistics Reform has been created with the aim of building EBPM (Evidence-based policy making), improving economic statistics to enhance GDP statistics, and addressing other issues faced by the statistical system of Japan.

In response to the Statistics Reform, the "Master Plan Concerning the Development of Official Statistics" (hereafter referred to as the Master Plan), which is currently in its third version, starting in April 2018 and lasting until March 2024, has been updated. The Master Plan was firstly created in March 2009 based on the Statistics Act in order to promote comprehensive and systematic measures for the development of official statistics. The third Master Plan covers issues relating to important official statistics, such as GDP statistics, Consumer Price Index (CPI), and the Business Register. The Statistics Commission, comprised of statistics academics and statistics users, and established in the Ministry of Internal Affairs and Communications in accordance with the Statistics Act, is tasked with the deliberation of technical detail and the implementation of the Master Plan. In addition, as in the Statistics Reform, the Statistics Commission has had its function strengthened from the perspective of improving autonomy, mobility, expertise, impartiality, and neutrality (e.g., develop function of proposals, recommendations and follow-up). The Director-General of the Research and Statistics Department of the Bank regularly participates in the Statistics Commission in a personal capacity as a representative user of statistics and takes an active part in the discussion.

4. Conclusion

This paper discussed the Bank of Japan's efforts to improve the quality of its statistics, with an emphasis on the importance of communication between compilers and users. From the position of statistics compiler, the Bank communicates with users through a public consultation process when making substantial revisions of the statistics it compiles. At the same time, from the position of statistics user, the Bank contributes to improvements in economic statistics by making comments based on its understanding of the statistics through its analysis work. An internal feedback loop between statistics compilers and users in the same department creates synergy effects. The Bank also contributes by participating in discussion regarding the government's statistics reform from a user perspective.

References

1. Bank of Japan (BOJ) [2009] "Bank of Japan's Basic Principles for Statistics: "The Basic Principles for the Compilation, Release, and Development of Statistics" and Present Priorities in the Development of Statistics." Outline of Statistics and Statistical Release Schedule.
2. BOJ [2016] "Planned of Revisions of the Tankan." BOJ Reports & Research Papers (in Japanese).
3. BOJ [2018] " Basic Policy of Rebasing the Services Producer Price Index to the Base Year 2015." BOJ Reports & Research Papers (in Japanese).
4. BOJ [2018] "Compilation Method of the "Wholesale Services Price Index" in Japan." BOJ Reports & Research Papers.
5. Kanafuji A., R. Mandokoro, N. Kato, and T. Sugo [2018] "Revision of the Consumption Activity Index to address the 2008 SNA and Improve Accuracy." BOJ Reports & Research Papers.
6. United Nations [2010] Indices of Distributive Trade: Handbook on Good Practices.
7. Yorozu T., A. Shibasaki, M. Osada, M. Higashi [2017] "Real Exports and Real Imports: Methodology and Tips for Analysis."

Stakeholder engagement throughout the statistical value chain – views from South Africa

Barend de Beer[1]
South African Reserve Bank

Abstract

The global financial crisis led to a critical reassessment of the role of central banks and their impact on domestic and international financial markets. This reassessment includes a wide range of functions, one of which is central banks' increased contribution to the compilation of macroeconomic statistics within national statistical frameworks. The demand for statistics from national statistical agencies and central banks has also increased with a renewed focus on the variety of macroeconomic statistics, methodological foundation and quality of statistical output. It has therefore become imperative for central banks to continually reassess their macroeconomic statistics compilation framework to meet these requirements and to maintain international best practice. An important component of this is the stakeholder engagement framework. This relates to the management of stakeholders throughout the value chain of the statistics compilation process from respondent to end users. Such a framework has three distinct but interrelated components. Firstly, communication with various respondents from which input data is sourced and secondly stakeholders within the compilation process. The third, relates to diversified communication strategies for the different user groups. The Economic Statistics Division of the South African Reserve Bank is in the midst of an assessment and extension of its macroeconomic statistics compilation function and the reshaping where necessary of its stakeholder engagement framework. This paper explores a possible statistics stakeholder engagement framework to address this component of the Banks' statistics compilation function.

Keywords

Statistics; National Statistics Framework; Stakeholder engagement framework; respondent

---

[1] The views expressed in this paper are those of the author and not that of the South African Reserve Bank.

1. Introduction

South African macroeconomic statistics is developed, produced and disseminated by Statistics South Africa which is the national statistical agency as well as the South African Reserve Bank[2], both operating under separate frameworks and governance. Stakeholders in the macroeconomic statistics framework include different groups such as academics, researchers, students, respondents, politicians, media, businesses, government and international organisations as well as the general public. These interest groups can be clustered as providers, producers, funders and users of information with each group having different expectations of the statistical agencies. Strategies to balance and respond to these expectations should account for the almost universally accepted premise that *change is constant*. There has been a notable shift in the macroeconomic statistics framework over the past two decades due to various factors such as the emergence of an increasing number of data sources, financial crises, and increased user demand which renders the frameworks with which statistical agencies compile statistics for the greater public good without too much concern for a multitude of stakeholders, dated. In the past, the concept of building a brand by agencies compiling macroeconomic statistics seemed foreign and not generally accepted as a requirement. However, with the advent of the 21st century and various concurrent changes, contemporaneous thinking now dictates that it is important to build a brand which can be associated with a value and output dictum. The brand should be visible, relatable, reliable and trustworthy while conveying the value proposition of the agency to its stakeholders.

2. Building a brand

The United Nations Economic Commission for Europe (UNECE) (2017) states that the "*Excessive modesty about official statistics is dangerous. Like other industries, we need not only to generate value but to demonstrate and publicise that we are doing so*." One of the initiatives of the UNECE has been the development of a set of recommendations regarding the advancement of the value of official statistics, thus focusing on the value proposition put forward by statistical agencies. Recommendation 6 deals with the development of a brand for statistical agencies. According to Sammut-Bonici (2015) a brand can be described as "*a set of tangible and intangible attributes designed to create awareness and identity, and to build the reputation of a product, service, person, place, or organization*". The UNECE recommends that

[2] Referred to hereafter as "The Bank", with specific reference to the Economic Statistics Division of the Economic Research and Statistics Department.

brand recognition should be seen as an explicit objective, which is based on the usefulness and quality of what statistical agencies does and this foundation should be augmented by specific tangible examples of the value of official statistics. Thus, in brand building, statistical agencies should: 1) Be more assertive about adherence to the Fundamental Principles of Official Statistics and the value that this generates; 2) Promote and publicize how official statistics around the world have added value; 3) Formulate and implement explicit brand recognition and promotion strategies; and 4) Take steps to cater for a wider and probably less informed range of users.

Brand building by various statistical agencies worldwide has increased, examples are Statistics Canada promoting their work with the slogan *"Serving Canada with high-quality statistical information that matters!",* and Statistics New Zealand with *"Unleash the power of data to change lives!"*. The Bank of Portugal (2017) contributes to branding awareness through publications such as *"Making sense of central bank data – The Banco de Portugal's experience in communicating statistics"*. The central hypothesis of the authors is that *"numbers alone do not tell a story – to mean something, their inner value must be brought to life"*. Branding should enhance the awareness of the various stakeholders as to the inherent value of the statistics. However, building valuable statistics is a dynamic process and to understand the value proposition requires an assessment of the statistical value chain (SVC).

3. Stakeholder engagement through the statistical value chain

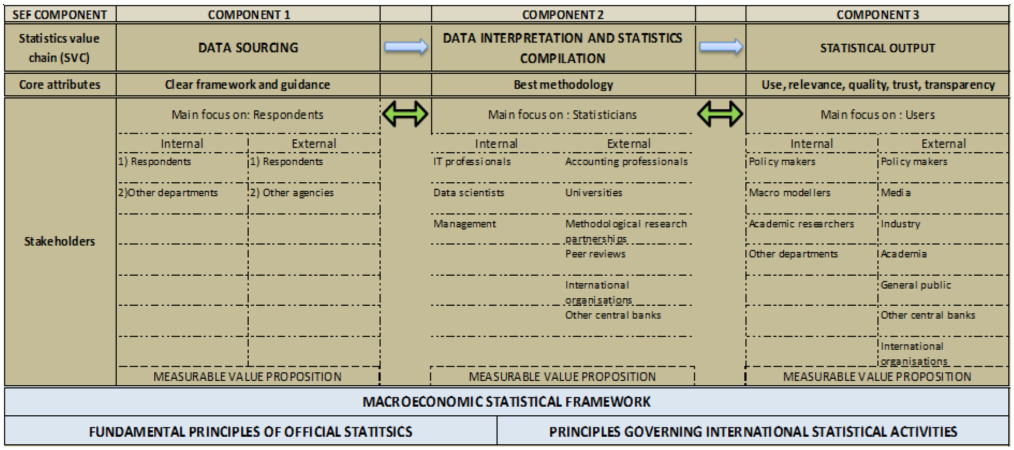
The macroeconomic statistical process is value chain driven with the ultimate goal of generating statistics that add value for the end user. Of paramount importance is the value proposition offered to users which requires an understanding of how value is added as the production process progresses. This poses the question: ***for whom do central bank statistical outfits build the value which is communicated through their value proposition?*** The conventional view would be that this is done to serve the end users of statistics. However, this is short-sighted as it only directs attention to one of the numerous stakeholders throughout the macroeconomic SVC. To correctly appropriate resources it should be acknowledged that end users only represent one of the stakeholder groups. While the previous statement is necessary, it is not sufficient in itself. Before engaging in the topic of what might serve as a sufficient stakeholder engagement framework (SEF), it is necessary to acknowledge that as a first order priority it is no longer the sole discretion of the Bank to decide whether it wants to engage with stakeholders or not. Given changing socio-dynamics in the first 19 years of the 21st century, active stakeholder engagement has become a prerequisite and a necessary condition. Once acknowledged, it requires decisions as to when and how to successfully engage stakeholders, an activity which is premised on the notion that *"those groups who can affect or are affected by the achievements of an organisation's purpose should be given the opportunity to comment and input into the development of decisions that affect them."* (Jeffery, 2009). It is thus more prudent to pre-emptively identify and actively seek out stakeholders in a managed approach, as opposed to haphazardly responding to their demands. However, in order for the Bank to manage its stakeholders it is necessary to understand their requirements so as to align the value proposition offered to their needs. This emphasises the need for a coherent and unified SEF.

Diagram 1 shows a possible visualisation of a SEF in the context of the Bank's macroeconomic statistical framework based on two foundational pillars, namely: 1) the fundamental principles of official statistics, and 2) the principles governing international statistical activities. The central question throughout the framework relates to what is understood as the value proposition rendered to each of the identified stakeholders, i.e. what is the value to them of the produced statistics? A key related question is how the value of produced statistics is measured, which is a topic that will be discussed later in this paper. At this juncture two related concepts – the SEF and the SVC, need to be merged, with the basic proposition being that the SEF needs to be built and managed around the SVC. In its simplest form the SVC has three distinct but interrelated components, namely; data sourcing, data interpretation and statistics compilation, and then statistical output. Each of the three components of the value chain needs to be assessed separately due to significant differences. By segmenting the SVC into these three components it allows to define the value proposition for each one and generate a SEF per component. Furthermore, the value proposition for each component needs to be built around practical and tangible criteria. Each of the three components will now be discussed separately.

Component 1 of the SEF relates to the stakeholders identified within the data sourcing phase of the SVC. The primary stakeholders are the respondents from which data is sourced. Respondents can be internal or external to the Bank and given their critical role in the data sourcing process it is non-negotiable to develop a stylised SEF aimed at managing the relationship with them. Tangible criteria that will form part of such an SEF can be summarised as follows: 1) Clear guidelines detailing expectations and providing guidance, 2) Dynamic survey forms with customizable parameters, 3) Optimised respondent burden throughout the national statistical framework (NSF), 4) Effective respondent support and engagement, and 5) Respondent self-help.

Diagram 1: Stakeholder engagement framework (SEF)

The author views the strategic pillars of Statistics Canada's respondent relations management programme as a good base from which to develop a fully-fledged respondent management framework. Although the respondent is identified as the main stakeholder, it is not necessarily the sole stakeholder. Strategic partnerships should also be formed with other stakeholders within the Bank as well as externally to augment/strengthen the ability to source higher quality input data. This could be due to various reasons such as the mandate that other departments within the Bank have which could support



the data sourcing effort. This type of relationship can similarly exist with external organisations in the NSF where mandates for data sourcing might be combined to reduce respondent burden and improve response rates. In either of these instances, the identified stakeholder should be managed as a key stakeholder within the first component SEF, which implies that they should understand the Bank's statistics mission, vision and brand as well as the clearly defined stakeholder value proposition. An alternative way of thinking of the respondent as a stakeholder is through the microdata approach to data collection. In this instance there is a joint commitment by the industry and the Bank to invest in the infrastructure for the compilation of granular statistics with micro data measurement structures on a voluntary basis in order to increase benefits and reduce costs in the medium to long-term.

The principle activity contained in the SVC for component 2 relates to the transformation of input data to compiled statistics and the complex and involved process this entails. In this regard statisticians need to identify internal as well as external stakeholders that contribute to this component of the SVC. Internally this would be the process of managing all stakeholders within the Bank that could make a contribution. Strategic partnerships would typically form a key element here and could for example entail involving other departments in the Bank to enhance the statistician's ability to do accurate and efficient computations. Given the reliance on Information Technology (IT) platforms and applications in modern day statistics compilation an important internal stakeholder is the IT department within the Bank. It becomes critical not only for the IT department to understand the IT requirements of statisticians but also the implications of IT constraints on the statistics value proposition. This presupposes a more refined bi-directional strategic partnership between the two functions of the Bank rather than just the more conventional one-directional relationship.

Other internal stakeholders that are separately identified for this component are data scientists and management. Although there is an overlap between the work of statisticians and data scientists there are also differences. Statistics is a mathematically orientated field that focuses on the collection, quantitative and qualitative interpretation of data and the generation of macroeconomic statistics while data science is multidisciplinary with a focus on scientific methods, processes, and systems to extract "knowledge" from data in a range of forms. Given the synergies between these two fields, this is a relatively recent but important partnership that has to be nurtured and catered for in component 2 of the SEF. The last explicitly identified relationship to be managed at this level is that of the executive management in the Bank. Executive management of modern central banks have numerous areas that require their attention. They inevitably acknowledge the importance of quality statistics, but the statistician should not take for granted their appreciation of the intricate requirements pertaining to the production of statistics. For statisticians to have adequate executive management support they should develop a dedicated stakeholder management programme. Externally, this component of the SEF could relate to institutions such as universities with whom strategic partnerships could be formed to develop statisticians. An excellent example of this is the collaboration between the central bank of Portugal and the Nova University of Lisbon which developed the Postgraduate programme in statistical systems with specialisation in central bank statistics. In this case the central bank and the university collaborate to provide formal training to central bank statisticians covering theoretical as well as practical modules in central bank statistics. This is one of the most vivid examples of managing stakeholder partnerships. In addition, the Bank could form methodological research partnerships with external organisations with project based research to enhance the skillset of the Bank's workforce without necessarily adding overall headcount. Other external stakeholders that should be noted here are peer reviewers that conduct assessments of the Banks methodologies and statistics output. Lastly, note should also be taken of international organisations with an interest in statistics, such as the IMF, OECD, BIS, IFC, UN, etc. as well as statistics departments of other central banks.

The third component of the SEF relates primarily to the end users of statistics as stakeholders. The diagram identifies different internal and external user stakeholders. Internal to the Bank the main stakeholders are policymakers, macro modellers, academic researchers, and other departments within the Bank. External end-user stakeholders are policy makers, media, industry, academia, general public, central banks, and international organisations. The value proposition to each of these identified stakeholders differs and requires a detailed plan to service their unique requirements. For statistics to be meaningful to the identified stakeholder segments they need to be properly channelled and communicated. In addition to the conventional channels to reach the identified stakeholders, the following are examples that can be explored to enhance delivery: 1) social networking platforms, 2) domain specific seminars and networks, 3) workshops targeted to industry groups, 4) interviews with newspapers and television stations with briefings focused on particular statistical issues, 5) regular sessions with journalists and media representatives, 5) short information videos on the web or television, 6) thematic sessions at universities and other training institutions, 7) stalls and posters at relevant events, 8) continuous refinement and adjustment of the monthly or quarterly statistics publications, and 9) metadata explanation in standardised non-technical language. In addition, to the above mentioned initiatives which could be implemented in the short to medium term, supplementary stakeholder initiatives for the medium to longer term are: 1) the hosting of specific methodological seminars exhibiting the methodological research done the Bank, 2) development of storytelling techniques to enhance delivery of statistics, 3) improvement and publication of dynamic non-stationary info-graphics with state of the art IT applications, 4) short and thematic statistical videos, and 5) development of a dedicated statistics portal. This range of tools can be stylised and customised to deliver and enhance the specific stakeholder value proposition in a customised fashion.

4. Measuring delivery on the stakeholder value proposition

After defining the stakeholders in the SVC within the SEF the success of the framework in delivering on the value proposition has to be assessed. This section draws significantly from the recommendations by the Economic Commission for Europe's task force on the value of official statistics (2017). Using those recommendations as foundation, a framework is proposed to measure the value of statistical outputs. This framework should contain the following three components:

- **Observable "objective" indicators:** Objective indicators to quantify the actual use of the produced statistics by the various stakeholders identified in the 3rd component of the SEF, which should be used in conjunction with indicators reflecting the adherence to the Fundamental Principles of Official Statistics related to the use, relevance, quality, and transparency of statistics.

- **Subjective" indicators derived from user satisfaction surveys:** Gauge the value of statistics in terms of the confidence and trust in the statistics reported by the stakeholders in the 3rd component of the SEF, focusing on the usefulness and accessibility of the compiled statistics. Use of dedicated user satisfaction surveys which could be conducted annually or bi-annually is recommended.
- **Methodologies to value/monetise the value of statistics:** Notwithstanding the usefulness of the first two methods of value proposition assessment, a 3rd and very powerful method would be the ability to put a monetary value on statistics or at least the impact of the compiled statistics. In order to measure this, the following two methods, amongst others, can be considered:
  - Cost based approaches: In this approach the cost of producing the statistics is calculated. One application of this approach is to infer that the value that a community is willing to pay for the production of statistics reflects the value it attributes to the statistics. Although this is a relatively straight forward method it has drawbacks as it does not take differences in productivity and quality into account when comparing data over time or across countries. This methodology also implies that if cost equals value then central banks could merely spend more on statistics to enhance the value obtained from it, which is not necessarily the case. It would however be advisable to compute such a value and also express it in current prices so as to gauge whether investment in statistics is growing in real terms.
  - Market (equivalent) pricing: In this method, the goal is to approximate the market price of the produced statistics by considering the market prices of similar products transacted in a competitive market and use that as a proxy.

Finally, it is important to know more about potential stakeholders who presently are not using statistics. Interest should not only lie in the current value of statistics, but also the greater potential value to non-users. This raises some obvious questions: 1) Why are these stakeholder groups not using produced statistics? 2) Is it because they are not aware of the available statistics? or 3) Is the value proposition not appealing enough or 4) Is the statistics not generated in the most appealing format or correctly timed? In addition to understanding the stakeholder groups that are not currently consuming the produced statistics, it would also serve the Bank well to improve its knowledge on what kind of statistics is required due to changing stakeholders' needs, in order to adjust the suite of statistics to adhere to these dynamic requirements.

5. Conclusion

Central bank statisticians by design fulfil a "service delivery" role in the NSF. This is rooted in the reason for their existence which is to make a visible and quantifiable contribution to the assessment of current macroeconomic state of affairs. Although their value proposition should be segmented to address their unique stakeholder base, their statistics outputs should strive to provide information that contributes to the solution of important contemporary issues within the economy. In pursuit of this it is important to have an accurate and focused view of the stakeholder to be served. This conclusion is echoed by the United Nations Statistics Division (2012) which states that "*a statistical agency has to understand its stakeholders and its users' needs. The statistical agencies should build and sustain very good relationships with all of their key stakeholders, including users, data providers, funding agencies, senior government officials, relevant community organizations, and the media*." An agency like Statistics Canada confers with this and further states (2016) that statistical agencies should consult stakeholders for the following reasons: 1) They contribute to identifying data gaps in statistical programmes, 2) They participate in consultations as new surveys are developed or existing surveys are improved, 3) They provide data, 4) They provide administrative data files that can be used to complement survey programmes, 5) They use data to inform programs and policies. Apart from better serving the needs of stakeholders, a "network" of stakeholders could also yield benefits to the Bank by providing insights and contacts that allows them to detect structural changes in the economy and distinguish them from issues which might be transitory. Given this, it is important that the Bank have a SEF, and in addition, it is also advisable to have a unit/team dedicated to interacting with stakeholders on a regular basis which is managed as part of the SEF. This team could potentially be housed within a data management unit which has primary responsibility for managing respondent stakeholders, but the overall SEF should have management involvement and report back channels to ensure uniform stakeholder management policies and to maintain strategic momentum.

References

1. Banco de Portugal, (2017), *"Making sense of central bank data – The Banco de Portugal's experience in communicating statistics"*, whole document.

2. Cadete de Matos, et al., (2015), *"Strategic coordination of a National Statistical System: the case of the Portuguese Statistical Council",* whole document.
3. Jeffrey, N., (2009), "Stakeholder engagement: Roadmap to a meaningful engagement", Doughty Centre, Corporate Responsibility.
4. Statistics Canada, (2016), "*Compendium of Management Practices for Statistical Organizations from Statistics Canada's International Statistical Fellowship Program*"
5. Sammut-Bonici, (2015), "*Brand and brand building*", January 2015
6. United Nations Economic Commission for Europe, (2017), "*Value of official statistics: Recommendations on promoting, measuring and communicating the value of official statistics"*

Internal communication with users at the Deutsche Bundesbank: Improving statistics and analytics

Stephan Müller
Deutsche Bundesbank

Abstract
Frictionless and well-attuned communication between those who compile statistics and those who use them is essential in order to bring the supply and demand of statistics into equilibrium. Central banks play a special role in this context, as they both compile and use statistics. For example, the Deutsche Bundesbank compiles monetary, financial, and external statistics, as well as a comprehensive set of indicators and seasonally adjusted economic data. This information system is used in the monetary policy decision-making process, as well as for macroeconomic and macroprudential analysis. The contact between economists and statisticians is of particular importance, as it helps statisticians understand the user perspective, and economists the statistical information and its limitations.
This paper describes different workstreams within the Bundesbank that have been established to facilitate the two-way communication processes mentioned above. This includes working groups, institutionalised forms of communication, and direct daily communication. The latter allows for the exchange of information at the operational level, which facilitates direct feedback. All forms of communication strengthen economic analysis and macroprudential policy on the one hand and provide valuable information enabling statisticians to improve their products on the other.

Keywords
Communication processes; user needs; decision-making process; valuable information; improved statistical products

1. Introduction

At central banks, communication processes between the producers and the users of statistical data play an important role. Internal users need high-quality data to fulfil their tasks and therefore need frictionless communication with the respective data providers. The Deutsche Bundesbank compiles in particular monetary, financial, and external statistics, and also comprehensive sets of indicators and seasonally adjusted economic data. These data are primarily used in the context of monetary policy decision-making as well as macroeconomic and macroprudential analysis. Well-attuned communication processes between statisticians and economists are thus helpful in gaining an understanding of the user perspective as well as the statistical information and its limitations.

This paper is structured as follows: following a brief overview of the overall tasks of Directorate General (DG) Statistics and how the user perspective can be taken into account, various types and levels of internal communication are illustrated by several examples. This starts with direct daily communication in everyday processes, especially at the working level, which facilitates direct feedback. Examples include microdata exchange and the work of the Bundesbank's data specialists. Particular emphasis is placed on the production process of the Bundesbank's Monthly Report. Another method of communication, for example, is the establishment of interdisciplinary working groups. This is mainly done in order to work on more conceptual issues. Finally, "institutionalised" forms of communication are explained. These mainly concern standardised processes for dealing with frequently recurring procedures. The final conclusion underlines that all forms of communication strengthen economic analysis and macroprudential policy on the one hand and provide valuable information enabling statisticians to improve their products on the other.

2. The objectives and producers of the Bundesbank's statistics

Statistics has been one of the fastest growing areas of activity at the Bundesbank for several years. The sharp increase in demand for statistical products due to macroprudential and microprudential requirements and the development of the Single Supervisory Mechanism (SSM) is causing a massive surge in subjects and users. In addition to monetary policy and the analysis of macrodata, the Bundesbank's DG Statistics also supports the areas of financial stability, markets, payments, banking supervision, and microdata research.

Specifically, the Deutsche Bundesbank is responsible for banking, capital market, and balance of payments statistics. For this reason, DG Statistics collects data from banks, non-financial corporations, and private persons. In addition, a large volume of statistical data comes from other institutions and is edited for analytical purposes. Statistics for the German economy are of particular interest. However, economic statistics for other countries are also becoming more important. A large number of indicators are seasonally adjusted.

The statistical results provided by the Bundesbank are used not only for internal purposes, but also serve as an indispensable basis for economic analyses in politics, business, science, and administration. The Bundesbank's statistics also represent the German contribution to the European statistical indicators provided by the European Central Bank (ECB) in order to fulfil the tasks of the European System of Central Banks (ESCB). In addition, the German results are also made available to other international organisations (e.g.

Eurostat, IMF, BIS, OECD) for the production of a wide range of international statistics.

3. Taking the user perspective into account

In terms of the Bundesbank's internal communication between the compilers and users of statistical products, questions are raised regarding the users' specific needs and focuses – the user perspective. Users of statistical products can be found, in particular, in the areas of monetary policy, economic research, financial stability, markets, payments, and banking supervision. They value the exchange of information with statisticians and actively promote it. Intensive contact usually takes place at the operational level. Users can make use of centralised functional email addresses belonging to the respective statistical departments for their requests. Subsequent in-depth discussions are usually carried out by telephone in direct contact between the user and the producer. Relationships are also fostered through informal communication in the form of joint lunches or other personal meetings. Users are especially interested in new developments in the statistical survey methodology, time series, or emerging technologies. In order to introduce users to these types of topics, personal contact is particularly important in order to remove ambiguity, become acquainted with the right contact persons, and gain a better understanding of the data and its relevance.

Regular cooperation also takes place in the production of joint reports or articles, for example in the context of the Bundesbank's Monthly Report. However, even texts written only by economists can convey new developments from the world of statistics, provided that these topics have been discussed with statisticians. Therefore, information about such messages must be actively shared with the users. This includes information on new findings in international bodies such as the Irving Fisher Committee on Central Bank Statistics (IFC), the Committee on Monetary, Financial and Balance of Payments Statistics (CMFB), or the European Statistical Forum (ESF). In this context, the exchange of information works in both directions. Innovations and new insights from the various areas in which the users operate can thus also be shared with the statistical community by providing them via these bodies. Other key areas in which intensive communication is important from the user's perspective are the production processes of statistical questionnaires and surveys, the interpretation of statistical developments in the context of the overall economic situation, as well as the provision of statistical data from internal and external databases. Especially in the latter case, users often lack a precise overview of the available data.

4. Workstreams within the Bundesbank to facilitate communication processes

a) <u>Direct daily communication</u>

In order to improve the internal communication processes between the producers and users of statistics, direct personal interaction forms the basis of positive relationships between different departments and contributes to the development of mutual cooperation. At the Bundesbank, such direct daily communication processes take place in many different workstreams and across various organisational boundaries. One use case is the communication between data users and data experts for the purpose of exchanging microdata. Bilateral communication and output orientation have priority here. In each area, there are dedicated data experts who act as contact partners for microdata users, and a single point of contact is currently being set up for all requests for microdata. Users come from different sectors, mainly from the economics, banking supervision, financial stability, markets, and risk controlling departments. It is important that these users communicate which microdata they are looking for as precisely as possible. To enable the users to do so, the data experts are conducting a kind of roadshow at the Bundesbank to present their services and give an overview of the available microdata. In the context of the provision of microdata, primary communication takes place at two specific times. The first instance is when the user contacts the data expert and describes the type of microdata that they are looking for. At this time, the data expert can give an initial response with first indications regarding the request. More importantly, however, this is the time for dedicated queries on data formats or the actual use case to specify the data needs. The second instance of intense communication takes place after the analyst has received the microdata. At this stage, communication mainly concerns clarifying ambiguities and interpreting the microdata. Statisticians also benefit from this kind of communication, as it helps them to gain a different perspective and to detect possible flaws in the datasets, which are then rectified.

A similar project was the introduction of data specialists, who assist users in searching for the desired data in internal and external databases. In particular, they are experts for the databases of international organisations such as the BIS, IMF, OECD, Eurostat, ECB, UN, World Bank, etc. as well as for databases from private data providers. An essential way of communicating with data users is by carrying out customised training. Thus, data users not only become familiarised with the different databases and the various types of data, but also with a range of tools, programmes and search engines. It is also possible to contact the data specialists by telephone or email at any time and ask them to find specific data or time series that analysts or other users need for their work. A similar service is offered to users of the Bundesbank's real-time database. This database contains a large volume of macroeconomic data for Germany in chronological order and therefore allows the reconstruction of the available information at a specific time in the past. In this context, statisticians benefit especially from communication with the users because they can thus detect statistical anomalies and better interpret the data.

An example of established communication processes between statisticians and economists from DG Financial Stability is the exchange of transaction data, which is closely related to the introduction of the European Market Infrastructure Regulation (EMIR). In this context, regular meetings at working level between experts from DG Financial Stability and DG Statistics take place to share insights on transaction data. Experts also communicate using an intranet user group. The value added from the perspective of financial stability is, for example, better monitoring of the derivative markets and the behaviour of market participants. However, the transfer of knowledge and skills works in both ways. Data scientists have expertise in handling large and complex datasets, and they have the computational power at their disposal to process them. Financial stability experts can report the experience and economic insights gained through analysis back to data experts. This comprises, in particular, the detection of economic mechanisms in the data and their interpretation. It may also include detection of potential anomalies or discrepancies, which helps statisticians to produce better data and anticipate user needs.

Another regular workstream in which direct communication plays an important role is the preparation of the Bundesbank's Monthly Report1. The economic statements in the Monthly Report are fundamentally based on the analysis of statistical data. Therefore, close cooperation is needed, in particular between DG Economics and DG Statistics. Although DG Economics, in principal, has direct access to the statistical databases, there is still a lot of coordination required between the providers and users of the data in order to ensure their proper use.

Statisticians' main task in the production process of the Monthly Report is to review the included statistical data. The data must be checked for up-to-dateness, correctness, quality, and validity. At the same time, an examination of their confidentiality must be carried out. Although these tasks may sound boring, they are indeed very useful. It is also important to ensure that the data are not over-interpreted. Since there has been long-standing cooperation in this workstream, a number of routine processes can be used. This also applies to other regularly performed tasks, such as seasonal, working-day, or weather adjustment. As an example of working-day adjustment, the phenomenon of "bridge days" (working days that fall between public holidays and/or weekends) can be considered. Industrial output will be lower on these days, so this can be a reason for lower statistical figures. Statisticians recognise this effect and can clarify the statistical anomaly in this matter through their communication with economists. Statisticians also review the texts with regard to the respective formulations, but only in terms of reasonable statistical interpretation of the data. The assessment of the economic perspective is still solely based on the considerations of the economists.

In the event of special occurrences, statisticians can also provide their own individual contributions to articles in the Monthly Report. This provides an opportunity for a closer look at the statistical background of noteworthy economic events. These contributions benefit greatly from the communication relationships already described. An apposite example, where a well-established communication process was very useful in the preparation of the Monthly Report, is the section on the German economy in the February 2019 edition. Within this section, the statisticians provided a quantification of the effects of plant holidays and strikes on German motor vehicle production, which was processed in a separate box regarding the weakness of German motor vehicle production in the second half of 2018. It explicitly describes why the production of motor vehicles and motor vehicle parts fell by 8% on the quarter in seasonally adjusted terms in the third quarter of 2018 (Deutsche Bundesbank 2019).

The communication that takes place in support of the described production process of the Monthly Report has become largely standardised. This is necessary because the time frame for coordination between economists and statisticians is rather limited. For the statistical review, a time slot of two to three days is provided. Sometimes ad hoc coordination is required if changes need to be integrated into the Monthly Report at short notice. The communication takes places mostly via email. This includes personal email as well as coordination via centralised functional email addresses accessible to a designated group of employees. Mutual exchange via telephone is also possible, especially in the context of the last-minute changes described above. It is very important that, in this context, communication takes place almost exclusively at the operational level, which means that the experts themselves work closely together. However, communication across hierarchical levels is also possible, so that experts at the operational level can, if necessary, advise senior managers. Communication in and of itself depends to a large extent on employees actively seeking out interdisciplinary exchange and maintaining contacts independently. This is currently complicated by the fact that DG Economics and DG Statistics are not located at the same premises. Personal contact between the respective counterparts is therefore limited and other forms of communication must be used instead. The communication in the production process of the Monthly Report is typically based on the initiative of both sides. On the one hand, economists need data and ask statisticians to review the use of the data in the text. On the other hand, statisticians actively approach economists when they see potential difficulties in using the data. In this context, both sides look at the data from a different perspective. Economists are more likely to examine the macroeconomic situation from a macro perspective. By contrast, statisticians look more closely at the individual data and their potential impact on economic development.

When considering communication processes in the context of the Monthly Report, as described above, it is always important to note that the overall goal is to provide the reader with the most accurate information possible. This means that the reader's interest in economic developments must be at the focus. As a result, the text of the Monthly Report may need to be adjusted to avoid statistical difficulties and to provide a clearer picture of the situation that can better convey the intended messages. Close and personal cooperation at both the operational and managerial levels that emphasises the advisory nature of the role of statisticians thus helps to significantly improve the Monthly Report, which is one of the Bundesbank's key publications.

b) <u>Work on conceptual issues</u>

The implementation of temporary and interdisciplinary working groups provides a suitable framework for joint and cross-domain cooperation. In particular, more detailed analysis of conceptual issues is possible in this way. Communication between different working group members, who come from different subject areas, therefore takes place in a fairly organised way, but also regularly and directly between the different experts. A good example in this regard is the temporary working group of DG Economics and DG Statistics that was set up in early 2015 to investigate the effects of the relocation/re-establishment of national business units of multinational enterprises (MNEs) on national and international economic statistics, the possible problems in the collection of statistical data, and conceivable practical solutions. As part of the working group, about ten experts from DG Economics and DG Statistics have worked closely together on developing proposals for possible approaches on how to enhance the composition of GDP with regard to the effects of globalisation from the user's point of view.

The interdisciplinary cooperation and resulting communication promote a fundamental understanding of statistical concepts and practice among the users. For example, a better understanding of the economic ownership principle can be acquired. Conversely, this kind of communication can give statisticians an understanding of the users' basic needs. This entails familiarising statisticians with the possible economic uses of their statistical data. This could, for example, involve the use of GDP data for estimating potential growth and determining the output gap. In this way, the user requirements can be introduced to the relevant expert committees in order to further develop the respective statistics. In the case of the aforementioned working group on MNEs, the results were presented to the CMFB (Kirchner 2018). This offered an excellent opportunity to bring the user requirements that were discovered to a high-level committee. The relevant question with regard to meeting the users' needs was: "What should data tell us?" The open communication in the working group led to the corresponding answers: use and utilisation of domestic productive factors, changes in production capacity, imbalance in b.o.p. or unit labour costs, interdependencies between countries, and analysis of spillovers. The joint cooperation in such a working group also highlights the limitations of statistics. As the final results of the working group were published in a contribution to the Bundesbank's Monthly Report, the public was also informed about the potential, but also possible restrictions, of statistical work (Deutsche Bundesbank 2018).

DG Payments and Settlement Systems set up an internal exchange of ideas on the current developments in the areas of fintech and crypto assets. This format can be considered a kind of informal expert group with participants from DG Payments and Settlement Systems, DG Banking and Financial Supervision, DG Financial Stability, DG Markets, and DG Statistics. It is therefore not a working group in the true sense, but still provides a suitable framework for conceptual work on fintech-related issues. Here, too, interdisciplinary communication is key to success. In particular, the forum facilitates cooperation at the operational level. Each member of the group keeps all experts informed about current proceedings in their department. This allows each participant to gain an overview of new insights in order to disseminate them throughout the different areas of the bank that deal with fintech-related topics. The group also has a "living document", in which all new insights and current developments are registered. Thus, there is always an up-to-date document for tracking the latest findings. Statisticians can use the exchange of information to find out what is expected from them. At the same time, users can be informed about the range of statistical products. Since such platforms also ensure that statisticians and users become acquainted with each other, they also learn what their colleagues are currently working on. This therefore also helps to avoid duplication of work and can be understood as being linked to the statistical principle of collecting data only once.

c) <u>Institutionalised forms of communication</u>

Communication between the compilers and users of statistical information also works beyond direct face-to-face communication and personal contact. Of course, personal contact should be preferred where possible, as it offers the highest potential for improving statistics and analytics, as shown in the previous sections. Due to their organisational and physical separation as well as the sheer size of the various departments, however, the responsible contact persons are sometimes unknown. For such cases, most of the DGs of the Bundesbank have set up specific administrative units that can serve as central contact points and forward enquiries to the relevant experts. This applies to requests from users to producers, but also vice versa. From a statistics point of view, this is especially true for enquiries where the exact contact person is not clearly known or many units in all of DG Statistics are affected. An example here is the central email address for coordination issues in DG Statistics. The staff members who manage this contact point are familiar with the detailed organisational structure of DG Statistics and can act as an interface between the producers und users of statistical data. They can request input from the appropriate bodies and communicate with the respective statistical experts. Subsequently, the information provided can be consolidated and made available to the users. If it is possible and offers benefits, direct contact between the producer and the user can also be established. All in all, the use of a one-stop shop offers several advantages, namely a quick coordination process, the use of accumulated expertise, compilation of accurate answers, and the fact that users have access to a contact person at all times.

5. Discussion and conclusion

The discussion of the various examples of internal communication processes between the producers and users of statistical data at the Bundesbank – whether a daily direct or a more institutionalised form of communication – has shown its enormous importance for the relationships between the different stakeholders and its massive influence on the quality of economic analysis and statistical products. It was also shown that economists can gain a better sense of statistics because communication helps them to reconcile their picture of the economic situation with the "reality of the data". In this context, close cooperation with personal contact, especially at the operational level, and emphasis on the advisory nature of the role of statisticians is of particular importance. A fruitful exchange hinges on well-established personal contacts and willingness to think in interdisciplinary terms and across departmental boundaries as the process follows people, not the other way round. This close relationship promotes mutual understanding and sensitises both sides to their respective needs.

As a result, statisticians gain a more analytical view of their own results. This allows for higher accuracy in statistical quality control and is helpful in developing statistics further. As a consequence, the credibility of statistics, and thus the confidence in statistical statements, is increased. Ultimately, even in times of digitalisation and communication via electronic media, nothing compares to personal relationships and direct communication with peer-to-peer exchange of information.

References

1. Deutsche Bundesbank (2018), Activities of multinational enterprise groups and national economic statistics, Monthly Report, October 2018, pp 65-79.
2. Deutsche Bundesbank (2019), The weakness of German motor vehicle production in the second half of 2018, Monthly Report, February 2019, pp 47-48.
3. Kirchner, R. (2018), Multinational Enterprises: What should be measured for what purpose? Paper presented at the CMFB Workshop on Globalisation, Vienna, July 2018.

Salary EFT indexation: Preliminary methodological considerations

Wian Boonzaaier
South African Reserve Bank, South Africa

Abstract
Alternative or non-traditional data sources are increasingly being explored to supplement or validate survey based statistics. This area of research is especially relevant in the context of monetary policy in South Africa, where estimates for wage growth derived from survey based publications have produced conflicting outcomes over the short- to medium term, even when controlling for methodological differences. This has implications for the inflation forecast and interest rate decisions. This paper builds the methodological foundations needed to construct an annually chain-linked monthly salary index based on daily interbank electronic funds transfers derived from an automated clearing house of interbank retail payments in the payment system of South Africa. Ultimately, the index could be used to supplement survey based measures of wage growth, and more broadly to validate existing monetary and financial statistics.

Keywords
Electronic funds transfer; national payment system; remuneration; bilateral price index

1. Introduction

The Economic Statistics Division (ESD), situated within the Economic Research and Statistics Department (ERSD) of the South African Reserve Bank (SARB), is responsible for addressing a large portion of the recommendations put forward via the Group of Twenty (G-20) Data Gaps Initiative (DGI), which is currently in its second implementation phase. This has led to an increased drive to create or uncover new data sources which will assist in complying with the aforementioned recommendations, including validation of existing statistical production series.

As a result, the division has engaged BankservAfrica, an automated clearing house in the National Payment System (NPS), to explore the potential of the data underlying its electronic funds transfer (EFT) service. This project gained further impetus based on feedback from a number of Monetary Policy Committee (MPC) meetings where the issue was raised that selected indicators of wage pressure in the economy produced inconsistent growth outcomes in the short- to medium-term. It was therefore requested that alternative sources of data be investigated in order to validate existing labour market survey outcomes.

The objective of this paper is therefore to highlight the importance of interbank EFT transactions from the NPS as an alternative source of information for research on selected labour market dimensions, and more generally to augment the economic statistics compilation process within ESD. It will attempt to achieve this objective by addressing the key methodological considerations associated with constructing a chain-linked monthly salary index using retail interbank EFT data for the period July 2017 until October 2018.

2. National Payment System in South Africa

a) Institutional Context

The NPS, which is legally governed by Section 10(1)(c)(i) of the amended South African Reserve Bank Act No. 90 of 1989 (SARB Act) and the National Payment System Act No. 78 of 1998 (NPS Act), provides an efficient process for parties wanting to engage in a transaction to exchange value in an efficient manner. The system encapsulates the overall payment process, and all dimensions associated with effecting a payment transaction (including the relevant systems, procedures, and agreements) are integrated into the NPS. The NPS framework can be explained in terms of various interlinked networks and systems (SARB 2008):

1. *Customer networks.* Customer networks typically include networks established by large business customers[1] of commercial banks, which enables them to participate in the payment system, and to offer payment services to clients.
2. *Payment networks.* Payment networks consist of system and communication solutions implemented by commercial banks to facilitate payments for their clients. This would include automated teller machines (ATMs), internet banking, and payment instruments such as EFTs, debit orders and credit cards.
3. *Clearing network*. Bank as well as non-bank participants are allowed in the clearing network. Agents known as payment clearing house (PCH) system operators (SOs) or clearing houses also participate in the clearing network and is legislated to clear on behalf of two or more participants in the settlement system. The clearing of retail[2] interbank payments in South Africa is done by BankservAfrica, which is owned by South African clearing and settlement banks.

[1] This includes major supermarket chains and public utilities.

[2] Retail payments include cheques, credit and debit card transactions, online payments and automated teller machine transactions (SARB 2008).

4. *Settlement networks.* The South African Multiple Option Settlement (SAMOS) system forms the foundation of the settlement system. Settlement system participants have an account at the SARB and are required to post collateral, which ensures liquidity and the effective functioning of the settlement system. Based on SARB (2018), the SADC Integrated Regional Settlement System (SIRESS) facilitates cross-border settlement among SADC countries, and is available to all participating banks. In terms of international settlements, the rand is accepted as currency within the continuous linked settlement (CLS) system, which provides settlement service across central banks, real-time gross settlement (RTGS) systems as well as payment netting systems.

The focus of this paper will be on the population and subsets of retail interbank EFT payments administered by BankservAfrica.

b) Remuneration Payments in South Africa

There are various pieces of legislation which separately governs certain aspects of remuneration payments (both cash and electronic) in South Africa which have important practical implications in terms of constructing a salary (or remuneration) index. According to the amended Basic Conditions of Employment Act No. 75 of 1997 (the BCE Act), the remuneration[3] of employees should adhere to the following requirements[4]: (1) It must be paid out in rand currency, (2) it must be paid on a daily, weekly, fortnightly or monthly basis, and (3) it must be paid in cash, cheque or direct deposit.

A related legislative requirement is that said remuneration should be paid within seven days of completion of the relevant work cycle or termination of employment. Pension and provident fund payouts are not subject to the seven day rule. Given that remuneration can be paid out in four different frequencies and should be done within seven days of the last day of a wage or salary period, transfers can technically occur on any day of the month. The procedure to follow for weekends and holidays are not explicitly treated in the BCE Act, although the majority of large employers make the necessary arrangements to pay out salaries and wages earlier in cases where the original pay date falls on a public holiday or weekend. Deductions that are legally required to be withheld by employers from employees' remuneration include income tax, unemployment insurance and emoluments attachment orders.

---

[3] The term "remuneration", as per the BCE Act, is defined as "...any payment in money or in kind, or both in money and in kind, made or owing to any person in return for that person working for any other person, including the State..."

[4] Members of the National Defense Force, National Intelligence Agency, the South African Secret Service and unpaid volunteers are not subject to the BCE Act.

3. Description of Data

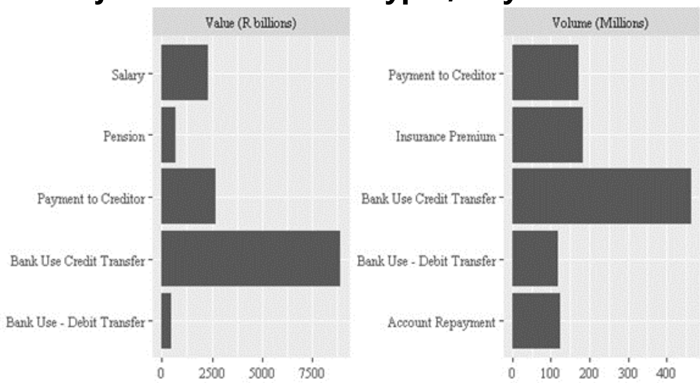
The data used in this paper is the population of all retail interbank EFT transactions on a daily frequency facilitated by BankservAfrica for the period[5] July 2017 to October 2018. The Structured Query Language (SQL) database is currently accessed online using *QlikView* business intelligence software. Same bank to same bank transactions are not covered. Interbank transactions are estimated to represent between 70 to 80 per cent of all EFT transactions, although it is not known how this proportion varies over time.

There are 54 fields that are populated for each EFT interbank transaction, some of which are only for internal use by BankservAfrica. The *Transaction Value* and *Transaction Volume* variables, for purposes of this paper, only include finalized transactions. Transactions which are subject to dispute or are classified as unpaid6 are not included. The *Transacting Bank* and *Transacting Bank Type* fields are typically analyzed together when looking at transactions. A bank is either the destination bank (referred to as the *homing* bank) or the originator (referred to as the *sponsoring* bank) of the transaction. The User Code is a unique identifier assigned to the client initiating the transaction.

There are 44 options available for classifying an EFT transaction, with a small portion of transactions bearing no classification. The top five EFT transaction reasons, measured in terms of value and volume over the indicated period, is shown in Figure 1. Credit transfers, which are payments initiated by a client to transfer some arbitrary amount to another account at another bank, represented more than 50 per cent (or approximately R8.9 trillion) of the total transaction value of all EFTs over the measured period (July 2017 to December 2018). Payments to creditors were the second largest category (slightly less than 20 per cent), followed by salary and pension EFT transactions. On the volume side, credit transfers remain the dominant driver, while insurance premiums also feature prominently.

[5] The short period made available through the online interface is most likely due to the high volume of data. 6 Disputed or unpaid transactions represented slightly more than 5 percent of all transactions in terms of volume and about 0.5 per cent as a portion of total value.

Figure 1: Major EFT Transaction Types, July 2017 – October 2018



4. Methodological Framework

The scope of the South African EFT salary index will be all monthly interbank EFT transactions that are classified as salary payments, net of all deductions made by the employer prior to effecting the payment, which are paid on the 15th, 25th or at the end of each month. The remaining salary payments (approximately 25 per cent of all salary transfers per month in value terms), including those classified as daily and weekly wage payments, will therefore not be considered for this index.

A price index, which will form the foundation of the methodological framework, can be described as "…a weighted average of the proportionate changes in the prices of a specified set of goods and services between two periods of time, say a reference period 0 and current period t." (SNA 2008). The focus will therefore be on fixed basket and chain-linked price indices. For purposes of calculating fixed weighted indices, the basket of EFT initiators in each period should be the same as in the base or reference period:

$$S \in S_0 \cap S_t,\ 0 < t \leq 15$$

where S is the matched basket of salary EFT transaction initiators and the base period was chosen to be July 2017. Chain-linking allows for the inclusion and exclusion of new and old items respectively:

$$S \in S_t \cap S_{t+1},\ 0 < t \leq 14$$

The EFT database consists of multiple daily transactions per EFT initiator per day. To obtain a price measure per EFT initiator per month, average transfer values per EFT initiator per transaction firstly need to be calculated as $\bar{p}_{ij} = \frac{v_{ij}}{q_{ij}}$. The variable $\bar{p}_{ij}$ represents the average price (or transfer value) for transaction j, effected by EFT initiator i, while v_{ij} and q_{ij} are the corresponding total value per transaction and the number of transfers per transaction. To obtain a daily average transfer value per EFT initiator, the harmonic mean of $\bar{p}_{ij}$ can be calculated as $\bar{p}_i^t = \left(\frac{\sum_{j=1}^{m}(\bar{p}_{ij})^{-1}}{m}\right)^{-1}$. The variable $\bar{p}_i^t$ represents the average transaction value per EFT initiator per day based on m transactions.

The range of average transaction values coupled with high heterogeneity in terms of the transaction volume distribution among EFT initiators lead to the conclusion that value or volume weighting schemes would not be appropriate for purposes of monthly aggregation. To therefore obtain the average transaction value per EFT initiator per month, the harmonic mean of the daily average transaction value per EFT initiator falling within a calendar month are calculated (not shown).

The focus of the methodological framework will be on the following price indices, both in chain-linked and unchained format: (1) Laspeyres, (2) Paasche, (3) Fisher, and (4) Törnqvist. The Laspeyres based salary EFT price index consisting of n EFT initiators can be written as follows (SNA 2008):

$$L_p^t = \sum_{i=1}^{n} \left(\frac{p_i^t}{p_i^0}\right) s_i^0 = \frac{\sum_{i=1}^{n} \left(\frac{p_i^t}{p_i^0}\right) p_i^0 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} \equiv \frac{\sum_{i=1}^{n} p_i^t q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}$$

where p_i^0 represents average transaction value per EFT initiator per month in the base period while $q_i{}^0$ represents the total number of transactions. To remain consistent with the literature, the price variable is displayed as p instead of $\bar{p}$. The variable $s_i{}^0$ represents the base period proportions. The corresponding Paasche price index can be expressed as:

$$P_p^t = \left[\sum_{i=1}^{n} \left(\frac{p_i^t}{p_i^0}\right)^{-1} s_i^t\right]^{-1} \equiv \frac{\sum_{i=1}^{n} p_i^t q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}$$

The Fisher index is calculated as the geometric mean of L_p^t and P_P^t, or $F_P^t = \sqrt{L_p^t \, P_P^t}$. The Törnqvist index is calculated as the geometric average of the price changes, with a weighting scheme equal to the average value shares in the two periods, or $T_P^t = \prod_{i=1}^{n} \left(\frac{p_i^t}{p_i^0}\right)^{(s_i^0 + s_i^t)/2}$.

A case can be made that the chain-linked versions of the above discussed index formulas would under certain circumstances be more appropriate for purposes of developing a salary EFT index. Goodridge (2007) states, for instance, that chain-linking allows for more frequent changes to the basket composition, while current period information can be compared with the previous period instead of some arbitrary base period. Chained versions of the Laspeyres, Paasche, Fisher and Törnqvist indices will therefore be tested on the current salary EFT sample.

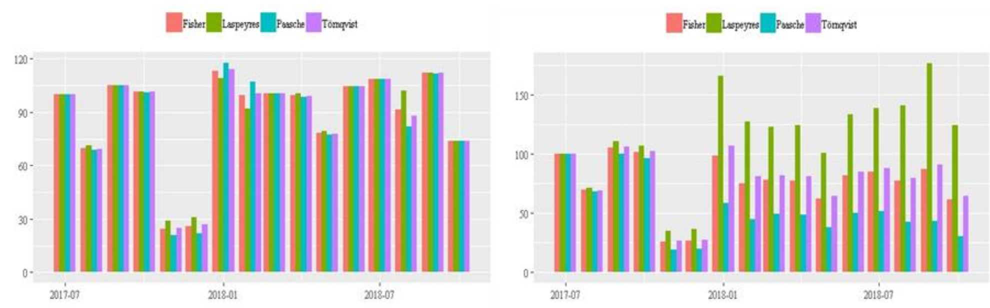
Outlier detection will be done on a quantile basis. The levels associated with minimum wages for domestic workers and an approximate threshold for high income earners based on the 2017 Tax Statistics will serve to guide the quantile cut-off values. For this purpose, only observations falling between the 20th and 90th percentile of the logarithmically transformed average transfer values per month will be included. No seasonal adjustment will be made given the limited sample size.

5. Discussion of Results

Figure 2(a) compares the unchained versions of the Laspeyres, Paasche, Fisher and Törnqvist indices for salary EFTs respectively. The reader's attention should immediately focus on the months of November and December in 2017. The steep decline in the index for the aforementioned months is primarily due to the strong seasonal increase in transaction volumes and average transaction values during November and December relative to July 2017, the base period used in the index calculation. This outcome supports the argument that a sufficiently long time series will be required to construct a statistically sound index. Following December 2017, the index returns to levels deemed to be in line with expectations.

Figure 2: Salary EFT Price Indices (July 2017 = 100)

Figure 2(b), which calculates the same indices as in Figure 2(a) but in chain-linked format, provides an even more interesting outcome. Apart from



the significant drop in the months of November 2017 and December 2017, the indices, with the exception of the Laspeyres index, do not revert back to "normal" levels. This can be viewed as an extreme case of *chain drift* and occurs in this magnitude when volatile movements (or bounces) are observed in the prices and/or quantities (Hill 2001). It is likely that the magnitude of the drift will reduce significantly once the index is constructed using seasonally adjusted data.

To reduce chain drift, the solution proposed by Hill (2001) is the use of *minimum spanning trees* to minimize the impact of the index number formula choice on the index outcome. The Paasche-Laspeyres spread (PLS) between time periods t and s, calculated as $PLS^{t,s} = \left|\log\frac{L_P^{t,s}}{P_P^{t,s}}\right|$, can be used to link periods with minimum dissimilarity. The similarity chained index in Figure 3 appears to resolve the problem of chain drift and does not differ much from the fixed base indices in Figure 2(a), although the methodology is statistically more robust.

Figure 3: PLS Adjusted Chain-Linked Salary EFT Price Indices (July 2017 = 100)

6. Conclusions and Recommendations

Based on the limited results in Section 5, it is recommended that an annually chained index on a monthly frequency based on the weights of the respective EFT initiators be tested for purposes of reducing the inherent volatility in salary EFT payments. An appropriate seasonal adjustment procedure which properly takes into consideration seasonality over multiple frequencies, the impact of public holidays and weekends, as well as other relevant regressors, should be applied to the data. The most appropriate outlier removal method should be based on an empirical study comparing various methods using the EFT database. A sufficiently long time series would be required to properly implement the above discussed recommendations.

There are some outstanding research questions. First, the issue of what constitutes "price" in the context of salary EFT indexing along with the most appropriate weighting scheme should be given careful consideration, as some economists may find the current use of average transaction value problematic. Second, the view that EFT initiators can be represented as a basket of items exhibiting different price and volume characteristics could be debated on a conceptual level. Third, it should be considered whether more than one index should be produced, given that the users of this index will have different needs and requirements. For example, certain users may be interested in a remuneration index consisting of daily, weekly as well as monthly salary payments, while others may be interested in a subset of payments.

References

1. BankservAfrica. 2018. “Take-Home Pay Reflects Massive Increase in August.” BankservAfrica. https://www.bankservafrica.com/BDSI/Mikes-BDSI-report.
2. EC, IMF, OECD, UN, and World Bank. 2008. System of National Accounts. New York.
3. Goodridge, Peter. 2007. “Methods Explained: Index Numbers.” ONS Economic & Labour Market Review 1 (3): 54–57.
4. Hill, Robert. 2001. “Measuring Inflation and Growth Using Spanning Trees.” International Economic Review 42 (1): 167–85.
5. SARB. 2008. “Overview of the South African National Payment System.” Financial Stability Review, September. South African Reserve Bank, 39–43.
6. SARB. 2018. “Review of the National Payment System Act 78 of 1998.” Policy Paper, September, 1– 60.

Modernising statistics: Communicating the wealth of statistics in a digital age

Per Nymand-Andersen[1]

Unless a policy - and the statistics underpinning it - can be explained, it will not be understood and the institution and policymaker carrying it out will lose credibility and trust.

Abstract

Independent statistics are fundamental if we are to present and compare the structures and dynamics of our societies. Safeguarding and communicating facts and figures and facilitating the use of good quality statistics for evidence-based policymaking contribute to sound and sustainable policies for the collective benefits of citizens. Trustworthy statistics have a welfare-enhancing effect and constitute a knowledge base that is fundamental to credible and sound decision-making.

This paper demonstrates the need for the statistics community to use statistics as a strategic asset and to proactively provide the professional user community with statistics and sound methodologies which are easily understandable, presentable and digitally reusable as part of its efforts to reach out to a wider professional and public audience.

Communicating statistics is a fundamental responsibility of those institutions engaged in official statistics, both to meet their *transparency* and *accountability* objectives, and also to crowd out lowquality statistics and data-feeding sentiment-based policy strategies. The use of low quality and sentiment-based information may well amplify and have contagion effects, not only on the lack of trust but also on the use of official statistics. Consequently, the reputational loss goes far beyond those institutions involved in providing official statistics.

Keywords

Communication; strategy; independence; quality; official statistics; trust

1. Introduction – the challenges of digital transformation

In today's open market economies and modern societies, there are ample and freely rendered private and public digital data sources of varying quality for both national and international policy use. The effect of digital

[1] European Central Bank, per.nymand@ecb.int. The views expressed in this paper are those of the author and do not necessarily reflect those of the European Central Bank. Further details of the author: https://www.linkedin.com/in/pernymand-andersen-81609913

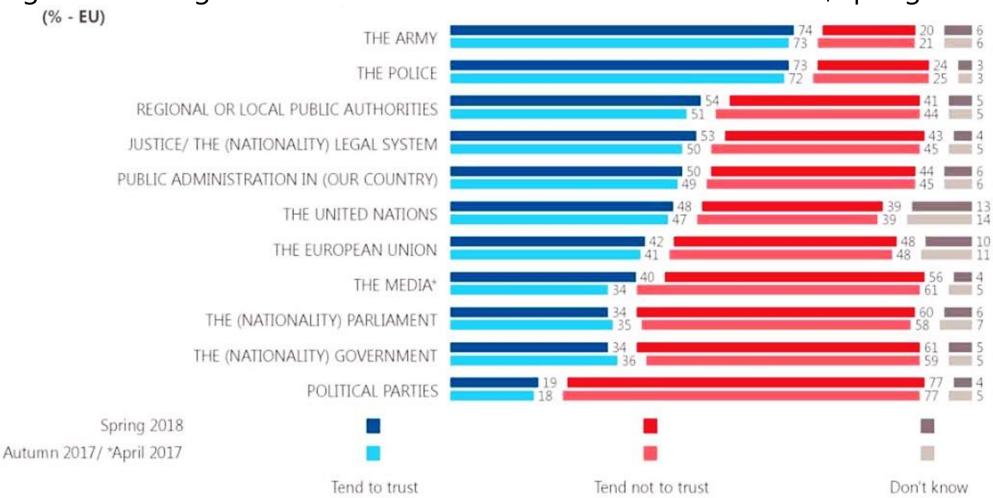
transformation in our society challenges the use of official statistics in multiple ways. First the *sheer volume* of available data/statistics and the *speed* at which they can be shared via digital channels for easy reuse and redistribution mean that official statistics are no longer the only data source available for users. Second, users need to be *source-critical* and spend the time necessary *to assess the quality* of multiple sources, which will enable them to differentiate between good-quality "official statistics" and other statistics of uncertain quality. Third, and perhaps more intuitively, professional intermediaries and users (e.g. policy analysts, advisers, journalists, opinion formers and policymakers) have less time to investigate *methodological differences* and their *impact* on comparability, policy analysis and any recommendations made as a result.

Furthermore, according to the European Commission's Standard Eurobarometer survey conducted in spring 2015,[2] an absolute majority of European citizens ***do not trust statistics*** and only 59% of respondents believe that statistical information plays a role in decision-making. The level of citizens' trust in national governments and parliaments is even lower; according to the Standard Eurobarometer survey of spring 2018[3] only 42%, 34% and 34% of respondents "tend to trust" the European Union, their national government and national parliament respectively. Moreover, 77% said they tend not to trust political parties.

[2] Eurobarometer (2015), Standard Eurobarometer survey on "Public opinion in the European Union", spring wave, No 83, European Commission, July. http://ec.europa.eu/public_opinion/index_en.htm. Unfortunately Eurobarometer has not continued to ask this question in subsequent surveys.

[3] Eurobarometer (2018), Standard Eurobarometer survey on "Public opinion in the European Union", spring wave, No 89, European Commission, June. http://ec.europa.eu/public_opinion/index_en.htm.

Figure 1: Average level of EU citizens' trust in certain institutions, spring 2018



Source: European Commission (Standard Eurobarometer survey)

These trends are critical, and, if not reversed, will have significant social consequences as citizens' lack of trust in public establishments, institutions and policymakers grows. It may well be that this situation has already materialised, since sentiment-based information seem to be increasingly – and deliberately – used within a political context, where tailor-made digital channels are used to reach large groups.

The sources and methodology of sentiment-based information are often opaque, and typically are deliberately obscured. Even attempts to debate the sources and methodology in a transparent manner seem not to be welcomed by those who use such information. This gives an impression of indifference to whether a description of reality is correct, as long as it suits the users' purposes.[4] It may even be a deliberate attempt to misrepresent reality without any accountability for the potential impact it may have on our societies. This is not a *new* phenomenon.

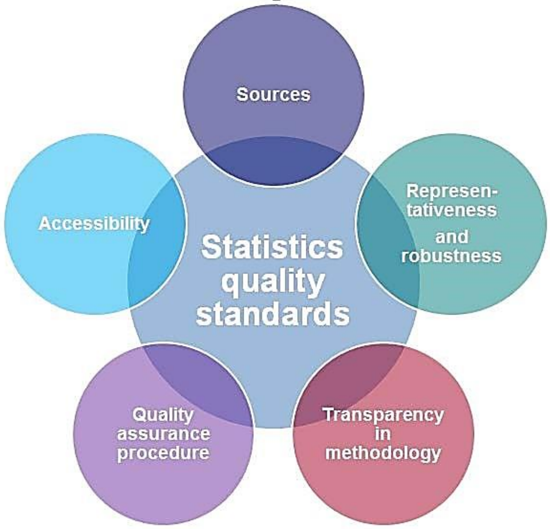
Official statistics, by contrast, involve measuring the structure and dynamics of our society. Wherever possible, they follow international statistical standards – which are available as a public good – in a transparent, objective and impartial manner.[5] Official statistics are of good quality and explain their

[4] Harford, T. (2016). "How politicians poisoned statistics" Financial Times, 14 April.
This article refers to a 1986 essay in the Raritan Quarterly Review by the Princeton University philosopher Harry Frankfurt defining the concept of and analysing misinformation in the context of communication.

[5] Official Statistics institutions and central banks publicly release their quality frameworks, which specify the main quality principles and elements guiding the production of statistics. The principles are generally based on those expressed in the Fundamental Principles of Official Statistics of the United Nations Statistical Commission and the related Principles Governing International Statistical Activities.

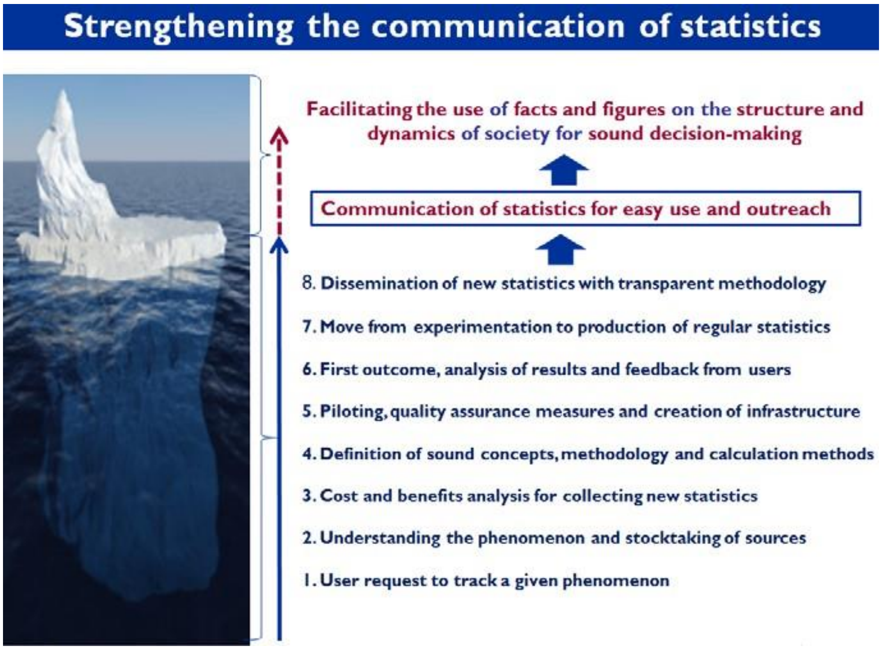
methodology and the underlying sources, while comments are invited to continuously enhance the final outcome.

Figure 2: Checklist for quality sources



Statistics agents may well play an important role in encouraging the use of good-quality statistics and also rebuilding trust. Their function is commonly focused on the core life-cycle of the statistical production process (see Figure 3), where they *implement sound methodologies* and *disseminate* official statistics in public databases. However, there is also a need to enhance the communication of statistics to address digital transformation and its challenges, by making statistics easily understood and facilitating digital reuse by various user groups.

Figure 3: The need to strengthen the communication of statistics.



This raises the following question: who has the legitimate responsibility to safeguard facts and figures and to facilitate the sharing and reuse of good quality statistics for evidence-based and sustainable policymaking?

2. Official statistics as trustworthy sources – taking up the communication challenge

The communication challenge is recognised by the international community. Eurostat and the European Statistical System have developed the **"European Statistical System Vision 2020"**, in which communication plays a new and important role. Communication needs to be conceptualised and operationalised by those agents responsible for providing official statistics and must demonstrate the importance of independent national and international official statistics in society.

The European central banking community has established a voluntary group – the "Statistics Accessibility and Presentation" Group (STAP) – to generate new proposals for enhancing the *accessibility and usability* of European System of Central Banks statistics and to lead and coordinate the implementation of the initiatives.

One of STAP's core deliverables has been he creation of the new *"Our statistics"* website[6], available in 23 official EU languages. The website aims to improve the communication of central banking statistics by presenting selected core statistics series using visuals and graphics, thereby facilitating

[6] See www.euro-area-statistics.org

the comparison and reuse of both euro area and national statistics. A core function of the website is the ability to easily share the statistics in digital form, either via email or social media, or by embedding the graphs into other websites and downloading data. The embedding capability is a particularly welcome feature for digital news media.

This website is innovative in many ways, and is based on a collaborative model, with functionalities shared with the OECD.[7] This is a concept can also be shared with national central banks or other agents. The codes are freely available reducing the time-to-market and costs. The website uses nonstatistical terms while also providing the corresponding statistical definitions. Furthermore, the website is accessible from PCs, tablets and other mobile devices, and also includes short statisticsbased narratives accompanied with visuals and tweets. It also uses open-source statistics software to monitor usage. In a similar vein, STAP is embarking on reusing (the codes) of one of Eurostat's interactive digital publications.

The key concept is keeping it short and simple, together with acknowledging that communicating understandable and easy-to-reuse statistics not only supports statistical, financial and economic literacy, but also facilitates the acceptance of sound and sustainable policy decisions. It is widely agreed that fostering transparency regarding policy decisions, by providing the underlying statistics, contributes to an efficient and effective pass-through for these decisions and also facilitates the external acceptance process. The broader the general understanding of statistics, the more likely it is that there will be support for sound and sustainable policies. The increased use of good quality statistics will lead to a better understanding of policy decisions and the proactive communication of official statistics will enable external users to use them appropriately, thereby providing their own analysis and contributions to public debate for the collective benefit of society.[8]

3. Visualisation of official statistics for user-centric communication

The modern world is a complicated place that faces increasingly complex social, financial, economic and political choices. This requires a more sophisticated way to communicate statistics. The challenge for the statistician is to extract knowledge from the large pool of available data/statistics and to present descriptive factual statistics in an economic and social context.

We must develop new techniques to manage, organise and release descriptive factual statistics that facilitate digital sharing and reuse, thereby adding user value by presenting them in a structured and interactive way. Here again, statistical organisations have a competitive advantage based on their

[7] See www.compareyourcountry.org

[8] See for instance, Blinder et al. (2008), Blattner et al (2008), Nymand-Andersen (2013).

long-held reputation for providing independent, factual and credible explanatory statistics of high quality.

Visualisation is a powerful digital tool that expands the statistical narrative. There are ample digital tools and freeware available to assist statisticians in presenting visual narratives such as interactive graphs, web-based videos, dynamic tables, infographics, touch-screen gadgets and smart mobile technology, which allow users to "slice and dice" relevant statistics and swiftly republish visualisations – without the need to find the data in large databases and/or even spend time to reproduce tables.

There is a growing gap between the way statistics are presented to society and the way the increasing digital community behaves, collects and reuses statistics.

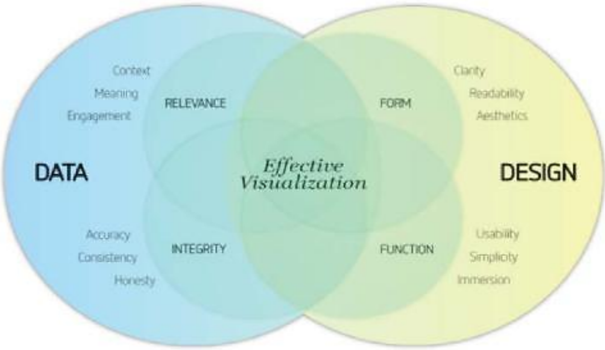
One simple and effective digital tool to share and cite statistics is the use of widgets, which can be easily incorporated into websites. In particular, the "embed" function allows professional users to reuse visualisation in other digital media/blogs and to write a short accompanying text.

Nevertheless, the concept of visualisation requires manifold skills and also collaboration between statisticians and other professionals such as designers and communications specialists. The statistician understands the narrative, the designer clarifies and creates the visual, and the communications specialist provides support in translating

> Research by the 3M Corporation in 2001 indicates that our brain can process visual information 60,000 times faster than text and that, on average, humans are more skilful at processing visuals, as approximately 90% of all information that comes to the brain is in that form (Hyerle, 2000). It has been suggested that human eyes can register 36,000 visual messages per hour (Jensen, 2008), while earlier research suggests that it takes approximately 0.25 seconds for the human brain to process and attach meaning to a symbol. In contrast, it takes the human eye approximately six seconds (a factor of 24) to read 20 to 25 words (Thorpe, Fize, and Marlot, 1996). While the research itself can be disputed, it nevertheless seems to support the idea that statistics presented in a visual format are easier to comprehend and retain than data shown as tables, numbers and words.

terms into non-technical language. We are competing for attention and people seem to spend less and less time on each topic, with the associated challenges for explaining complex phenomena in our society.

Figure 4: The statistical concept of digital visualisation – a multidisciplinary collaboration between statisticians, designers and communications specialists



Source: interactivethings.com

4. Conclusion

Producers of official statistics need to embrace digital and social media opportunities, as well as interacting and sharing communication concepts, methods and tools with other statistical organisations, also making them free for reuse. This is important as we are all engaged, both at national and international level, in making official statistics known, used and trusted – and crowding out the increasing volume of low-quality statistics and data.

Statisticians have an underutilised competitive advantage, as we already possess significant statistical knowledge and knowhow which can be adapted and shared with minimal effort. We have the benefit of international methodologies (standards), structured production facilities, similar data models, coding systems and data exchanges.

The challenge for the statistical community is to proactively extract relevant statistics and visualise these in order to explain and compare the structure and dynamics of our society in context, and also in a way that can be swiftly understood and reused in digital form by professional users as well as the general public.

The impact of sentiment-based policymaking and the price that citizens will pay for it greatly exceed the costs of official statistics to embrace the digital opportunities and build up a proactive user-centric statistical communication and publication strategy. But who else will safeguard official statistics?

References

1. 3M Corporation, 2001
2. Blinder, A.S., M. Ehrmann, M. Fratzscher, J. De Haan and D.- J. Jansen, (2008), “Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence”, *Journal of Economic Literature*, Vol. 46, No 4), pp. 910-945
3. Blattner, T., Catenaro, M., Ehrmann, M., Strauch, R. and Turunen, J. (2008), “The predictability of monetary policy”, *Occasional Paper Series*, No 83, ECB, Frankfurt am Main, MarcH
4. Eurobarometer (2015), Standard Eurobarometer survey on “Public opinion in the European Union”, spring wave, No 83, European Commission, July.
5. Eurobarometer (2018), 89, Standard Eurobarometer survey on “Public opinion in the European Union”, spring wave, No 89, European Commission, June.
6. Harford, T. (2016), “How politicians poisoned statistics”, *Financial Times*, 14 April.
7. Hyerle (2000), “A Field Guide to Using Visual”, Association for Supervision and Curriculum Development
8. Jensen, E (2008), second edition of Brain-Based Learning, the new paradigm of teaching. p 55. Corwin Press.
9. Nymand-Andersen, P. (2013), “Communicating central banking statistics: Making useful sense of statistics in a dynamic world”, *Statistical Journal of the International Association for Official Statistics*, Vol. 29, No 4, pp. 269-280.
10. Thorpe, S., Fize, D. and Marlot, C. (1996), “Speed of Processing in the Human Visual System”, *Nature*, 381(6582), pp. 520-522

French households' portfolio: The financial almost ideal demand system appraisal

Sanvi Avouyi-Dovi, Christian Pfister, Franck Sédillot
Banque de France

Abstract

French households have sharply increased their accumulation of financial wealth over the last three decades. This paper seeks to find explanatory factors of these changes. For this purpose, we estimate a financial AIDS (FAIDS) to French households' portfolio choices. This framework is an extension of the seminal AIDS model of Deaton and Muellbauer (1980a) to the financial portfolio.

We find that most of the estimated parameters of the benchmark model are in line with economic priors. In particular, wealth and real returns are the key determinants of the long run dynamics of the different shares. Moreover, some additional exogenous factors improve the general performance of the model, highlighting the role of uncertainty or demographic factors in the changes in the structure of French households' portfolio. The main results remain valid under some alternative assumptions implemented in the context of some robustness checks.

Keywords

Saving; model; wealth; return

1. Introduction

French households have sharply increased their accumulation of financial wealth over the last three decades. While their financial wealth represented less than 5 times their disposable income at the end of 1980, the ratio of wealth to disposable income reached around 11 at the end of 2016. In addition, the structure of households' financial portfolio greatly evolved over the period 1978-2016. For instance, currency plus overnight deposits and short-term saving deposits accounted for 58% of total financial wealth in 1978 and only around 30% in 2016. Life insurance contracts accounted for 4% of total financial wealth in 1978; about three decades later, this asset represents 44% of French households' portfolio. This paper seeks to find explanatory factors of these changes. For this purpose, the choice of a model structure has to be made.

In recent decades, many demand functions consistent with economic theory have been examined in the literature. Most of them have been described as flexible functional forms based on the duality theory. The most popular demand functions are the generalised Leontief functions (Diewert, 1971), the translog function (Christensen et al., 1975), the Rotterdam model (Theil, 1965 and 1975a and b) and the Almost Ideal Demand System (AIDS, Deaton and Muellbauer, 1980 a and b).

One of the advantages of both the AIDS and Rotterdam models is that they have recourse to theoretical restrictions that are statistically testable rather than impose them on the functional form. Barnett and Seck (2008) show that there is no definite evidence of the predominance of the AIDS models over the Rotterdam models in terms of explanatory power or forecasting performances. For example, the two classes of models perform well when substitution among goods is low. Also, according to their findings, the full non-linear AIDS models by Deaton and Muellbauer (1980 a and b) perform better than the original Rotterdam models in certain cases, whereas the Rotterdam ones clearly exceed the AIDS models in other cases. Furthermore, by comparing the AIDS and Rotterdam models using a model selection procedure (Amemiya, 1980, 1985), Erkan (2006) finds that the AIDS model displays better performances than the Rotterdam model.

As the aim of the paper is not to compare different demand systems, our selection of the demand system is guided by the best practices in this research area. Thus, the main guideline of our study is given by the robustness and stability of the empirical results from previous studies. As mentioned before, the AIDS models deliver relatively good performances under some conditions in many empirical studies. The appropriate framework should be clearly micro-founded in order to avoid some spurious relations. Finally, it should be empirically tractable and intuitively understandable. The AIDS model meets these requirements and is a good tradeoff between the different classes of demand systems.

In this paper, following Blake (2004), we apply a financial AIDS (FAIDS) framework to French households' portfolio choices. This framework is an extension of the seminal AIDS model of Deaton and Muellbauer (1980) to the financial portfolio. Such a model is compatible with the analysis of households' portfolio choices based on the neoclassical demand theory. In particular, the analysis of cross-interest rate elasticities enables us to assess substitution or complementarity effects between the different asset shares in households' portfolios.

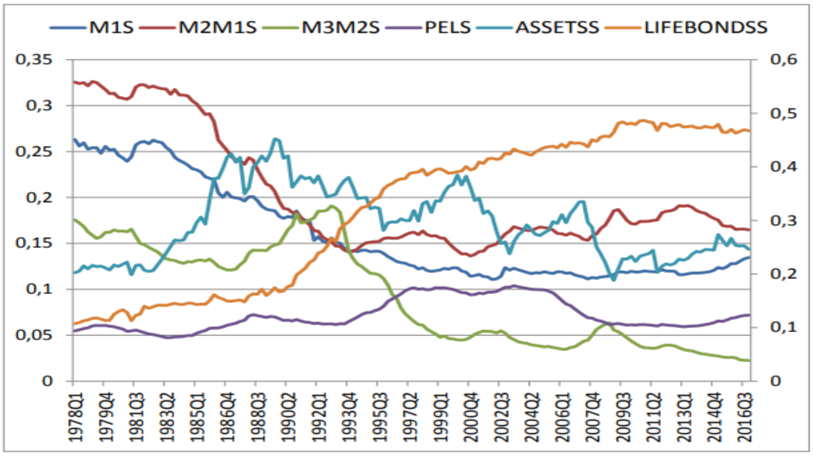
2. Methodology

We aggregate the individual financial assets to distinguish six categories of assets within French households' total financial wealth (Banque de France, 2005; Bachellerie et al., 2016). Thus, the share of asset "i, i=1,..., 6" represents the relative weight of this asset in total financial wealth (M1, M2M1, M3M2, PEL, ASSETS, LIFEBONDS). The six shares are the following (Chart 1):

- Currency plus overnight deposits, labelled "M1";
- Short-term savings accounts and deposits with an agreed maturity up to 2 years, labelled "M2M1";
- Money market fund shares, labelled "M3M2";
- Home savings plans (Plans d'épargne logement, PEL), labelled "PEL";
- Equities and non-money market fund shares, labelled "Assets ";
- The sum of life insurance contracts, long-term savings plans in securities distributed by banks (Plan d'Épargne Populaire, PEP), and debt securities held directly by households, labelled "Lifebonds".

For M1, we use the apparent interest rate on deposits that is available in the financial accounts for the period 2003-2016. We compute a weighted-average return for M2M1 based on interest rates for the different short-term deposits included in it. For M3M2, we use the overnight rate. For PEL, we use the interest rate set by government. For Assets, we compute the annualised return on the CAC40 stock market index. As we do not have access to returns on life-insurance contracts on a quarterly basis for the entire sample, the return on Lifebonds is the 10-year yield on Treasuries. Finally, to obtain the real returns, all the previous nominal returns are deflated by the inflation rate drawn from the CPI. Some additional explanatory factors are also tested in the relationships describing the dynamics of the shares

Chart 1 : Shares of Financial assets



Sources: Banque de France and authors' calculations. All shares except Lifebonds (LHS), Lifebonds (RHS)

3. Results

There are six shares but the baseline model is a dynamic system of only five equations. In order to meet adding-up conditions requirements conditions imposed by demand theory (Blake, 2004), we we drop one equation (the share of PEL, which is relatively weak) from the dynamics system and infer its parameters from the restrictions conditions. We add two exogenous variables (uncertainty or risk factor and demographic factor). The system is estimated with three-stage least squares over the sample 1999-2016. The set of instruments variables contains the lagged values of the endogenous and exogenous variables. Elasticities corresponding to wealth, interest rates and additional exogenous variables are derived from the estimates. The Student-t statistics are heteroscedasticity and auto-correlation consistent. The conventional validation tests (Wald, ADF, autocorrelation, normality, Arch, etc.) are also conducted.

Table 1 displays the results of the estimates of the coefficients of the fully constrained model with two exogenous variables. As regards the effects of the interest rates, we note that each asset is significantly and positively correlated with its own return. In the end all things being equal, an increase in the real return on asset “i” (i= M1, M2M1, ..., Lifebonds) leads to a rise in asset “i”. The sensitivity of the share to the real return depends on the asset, with the reactions of M1, M2M1 and Lifebonds to their own real returns being stronger than those of M3M2, Asset and PEL. However, whatever the asset, the uncompensated elasticity is significantly different from zero. One can also note cross-effects between real returns and financial assets. In particular, overall, the real interest rate of M1 negatively impacts the other shares; in this case, a substitution effect prevails. A similar pattern exists regarding the effect of the real interest rate of M2 on the other shares. In fact, overall, substitution effects prevail.

As regards wealth elasticity, the empirical results bring to light three situations. In the first case, there is a strong positive effect of wealth on Lifebonds. Only the richer households seem to react to the variations in Lifebonds. In the second case, we note a negative link between some financial assets (M1, M3M2) and wealth. According to Blake (2004), these assets are wealth-inferior assets with a long run negative elasticity. It is worth noting that these assets are relatively liquid. Maybe the changes in wealth are offset by the opposite variations in these liquid assets. In the last case, we note a weak correlation between two assets (M2M1 and Assets) and wealth as their coefficients in the corresponding regressions are not significantly different from 0. In particular, Assets seem to be substantially driven by their own returns and the substitution effects induced by Lifebonds.

Finally, only M1 is significantly correlated with the business cycle (through the de-trended unemployment rate) and financial uncertainty (via financial market volatility). Unexpectedly, the other shares, especially the Assets one, are not significantly impacted by real or financial risk factors.

Table 1: Estimates of the coefficients of the baseline model

	M1	M2M1	M3M2	PEL	Assets	Lifebonds
Wealth elasticity	-0.1 (-4.3)	0.0 (-0.8)	-0.1 (-2.6)	0.0 -	-0.0 (-1.2)	0.2 (7.9)
Uncompensated interest rates M1 real rate	6.0 (7.9)	-1.5 -	0.3 -	-1.5 -	-0.2 -	-3.0 -
M2 real rate	-1.5 (-2.1)	5.4 (5.6)	-2.7 -	-0.0 -	-1.3 -	0.1 -
M3 real rate	0.3 (0.6)	-2.7 (-5.0)	2.6 (3.5)	-0.0 -	-0.9 -	0.8 -
Pel real rate	-1.5 -	-0.0 -	-0.0 -	0.5 -	1.4 -	-0.3 -
Assets real rate	-0.2 (-0.6)	-1.3 (-3.3)	-0.9 (-2.0)	1.4	3.3 (4.2)	-2.3 -
Lifebonds real rate	-3.0 (-6.5)	0.1 (0.1)	0.8 (1.1)	-0.3 -	-2.3 (-3.5)	4.8 (5.1)
Detrended unemployment rate	-0.8 (-2.9)	0.4 (0.9)	-0.4 (-0.7)	0.0 -	-0.6 (-0.6)	0.5 (0.6)
Financial market volatility	0.0 (2.8)	-0.0 (-0.0)	-0.0 (-0.4)	0.0 -	0.0 (0.1)	0.0 (0.3)

We use the empirical model to simulate the effect on French households' portfolio allocation of the replacement of the various tax regimes of most financial products with a flat tax (FT) on savings income in 2018. Furthermore, the effects of a scenario, which we label "Extended FT" and in which the FT would be extended to all financial products (Pfister, 2018), are also shown.

Table 2 displays the long-term impact on outstanding amounts at end-2017, in variations from the baseline scenario, in percentage points of households' financial wealth and in corresponding amounts. In the long run, the FT would lead to a substitution out of Lifebonds and, to a lesser extent, M2M1 and M3M2 into Assets and, to a lesser extent, M1, with PEL being unaffected. Indeed, the FT would increase average taxation and thus lower the after tax return on Lifebonds and PEL (more precisely on life insurance contracts) and lower it on Assets (more precisely shares). Furthermore, in the baseline model, Assets are substitutable for Lifebonds (see Lifebonds equation) and complementary to PEL (see PEL equation), the latter effect offsetting the negative impact of the FT on the PEL yield.

An "Extended FT" would also cover so-called A passbooks included in M2M1. The column "Extended FT" in Table 11 shows that the substitutions resulting from the FT as it stands (into Assets and M1 and out of other liquid assets) would be strengthened if the FT were extended to all financial products' income.

Table 10: Impact of a flat tax on savings income
(Deviation from baseline, in percentage points and EUR billions)

	FT		Extended FT	
	Share	€ Amount	Share	€ Amount
M1	0,2	7,3	0,3	10,9
M2M1	-0,1	-4,4	-0,4	-16,0
M3M2	-0,2	-7,9	-0,1	-1,8
PEL	0,0	-1,4	-0,1	-1,9
Assets	0,8	31,5	0,9	34,0
Lifebonds	-0,7	-25,2	-0,7	-25,2

4. Discussion and Conclusion

Our model validates the FAIDS model as an appropriate framework for satisfactorily describing the dynamics of French households' portfolio. Indeed, this model highlights and quantifies the effects of the core explanatory variables (returns on assets, wealth) but also the roles of some additional exogenous factors. In particular, the model distinguishes the substitution and complementarity effects between financial assets, highlighting the role of wealth as one of the main drivers of the dynamics of certain financial assets and underlining the relative importance of income risk in the dynamics of these assets.

It is worth noting that even though the preliminary tests detected a break in the dynamics of the shares during the global financial crisis, the quality of the regressions over the period including the early 2010s remains acceptable.

This shows that the financial crisis did not lead to a major shift in French households' portfolio allocation.

The assessment of the effects of the flat tax, which has replaced the various tax regimes of most financial products in France, based on our model, shows a substitution out of life insurance contracts and, to a lesser extent, passbooks and other monetary assets mostly into equities, with home savings plans being unaffected.

As an avenue for future research, this study could be extended, for example, by introducing housing and debt into French households' wealth.

References

1. Amemiya T. 1980. Selection of Regressors, International Economic Review, 21 (2), 331-354.
2. Amemiya T., 1985. Advanced Econometrics, Cambridge: Harvard University Press.
3. Barnett W.A., Seck O. 2008. Rotterdam Model versus Almost Ideal Demand System: Will The Best Specification Please Stand up, Journal of Applied Econometrics, 23, 795-824.
4. Blake D. 2004. Modelling the Composition of Personal Sector Wealth in the UK, Applied Financial Economics, 14, 611-630.
5. Christensen L.R., Jorgenson D.W., Lau LJ. 1975. Transcendental Logarithmic Utility Functions, American Economic Review, 65, 367-383..
6. Deaton A, Muellbauer J. 1980a. An Almost Ideal Demand System. American Economic Review, 70, 312-326.
7. Deaton A., Muellbauer J. 1980b. Economics and Consumer Behavior, Cambridge University Press: Cambridge, UK.
8. Diewert W.E. 1971. An Application of the Shephard Duality Theorem: a Generalized Leontief Production Function, Journal of Political Economy, 79, 461-507.
9. Erkan E. 2006. Demand Systems for Agricultural Products in OECD Countries, Applied Economics Letters, 13, 163-69.
10. Pfister C. 2018. Fiscalité de l'Épargne et Choix de Portefeuille des Ménages Français, Revue d'économie financière, 131, 61-75.
11. Theil H. 1965. The Information Approach to Demand Analysis, Econometrica, 33, 67-87.
12. Theil H. 1975a. Theory and Measurement of Consumer Demand, Vol. 1. North-Holland: Amsterdam.
13. Theil H. 1975b. Theory and Measurement of Consumer Demand, Vol. 2. North-Holland: Amsterdam.

International comparison of the BIS international banking statistics: Estimates of the positions excluding trustee business[1]

Kimiaki Shinozaki, Kaori Saito, Kazunori Hiyama, Kyosuke Shiotani
Bank of Japan, Tokyo, Japan

Abstract
The BIS international banking statistics (IBS) are statistics that the Bank for International Settlements (BIS) and central banks worldwide compile in order to capture capital flows and credit exposures through internationally active banks. It is worth noting that the IBS reported by the Bank of Japan (BOJ) include trust accounts in which a bank holds and manages funds or securities on behalf of third parties. Attention should therefore be paid for international comparative analysis of the Japan's statistics and those of other major reporting countries which exclude trustee business. This paper attempts to provide an estimate of the IBS in Japan which exclude trust-account-based positions, and compare the results with the IBS of banks of other major nationalities. The results show that, while Japanese banks have been increasing their foreign claims in recent years, the ratio of foreign claims to the outstanding amount of total assets is not significantly high compared to banks of other major nationalities. Moreover, claims vis-à-vis the United States are demonstrated to account for a relatively high share of foreign claims, which lead to a high share of U.S. dollar-denominated claims held by Japanese banks.

Keywords
International Banking Statistics; foreign claims; cross-currency funding; trustee positions

1. Introduction[2]

The BIS International banking statistics (IBS) are global statistics that the Bank for International Settlements (BIS) compiles and publishes[3] in cooperation with the central banks worldwide in order to capture capital flows and credit exposures, comprehensively, through internationally active banks. While the Bank of Japan (BOJ) has been cooperating in compiling the statistics by providing the figures of the reporting banks located in Japan with the BIS quarterly, the BOJ has also been independently aggregating the figures and publishing the results (the IBS in Japan) in both Japanese and English.[4] The IBS are comprised of the following two datasets: (1) the Locational Banking Statistics (LBS), which measure flows of external claims and liabilities held by banks located in each reporting country, and (2) the Consolidated Banking Statistics (CBS), which capture international credit exposures of banks headquartered in each reporting country. The LBS and CBS enable one to identify capital flows in detail by merging these sets of statistics, such as positional breakdowns by reporting region and country, currency, and counterparty sector. Reflecting this point, they are used widely for analysis on topics related to international finance.

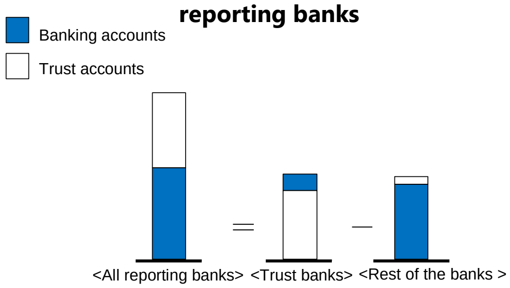
Capital and credit transactions through banks are accounted as either of the following bases: (1) banking accounts in which a bank's own funds are involved in transactions for purposes such as funding or fund management, or (2) trust accounts in which a bank holds and manages funds on behalf of third parties. In the IBS, reporting practices of trustee business vary across reporting countries. While many countries including major ones report figures of positions based solely on banking accounts, there are several countries whose reporting figures include positions based on trust accounts in addition to those based on banking accounts (banking-and-trust-account basis), such as Japan.[5] Differences in their coverage should therefore be taken into account when carrying out international comparative analysis using the IBS in Japan, notwithstanding their feasibility of capturing international capital flows and credit exposures through the banking sector in a broader sense, reflecting their wide coverage of funds and credit transactions.

In this regard, it is not currently possible to exclude trust-account-based positions from the IBS in Japan and compile positions based exclusively on banking accounts in order to conduct international comparative analysis. This limitation arises due to reasons related to the non-availability of breakdowns of reported positions. With the objective of carrying out an international comparison of the IBS based on banking accounts, this paper attempts to estimate the IBS in Japan wherein the positions are separated into those pertaining to banking or trust accounts under certain assumptions. We then provide comparative analysis of the estimated banking-account-based IBS in Japan with the IBS of banks of other major nationalities. Further, this paper outlines the characteristics of international capital flows and credit exposures held by Japanese banks through transactions undertaken on a banking-account basis.

2. Estimates of the positions on banking-account and trust-account bases

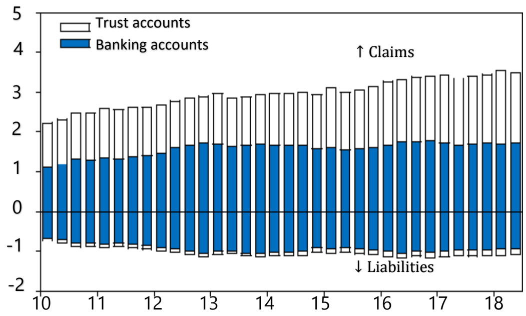
In the LBS, debt securities account for 74 percent of the external claims held by Japanese banks as of end-June 2018. While a country breakdown or a sectoral breakdown of debt securities held by individual reporting banks is not available for reasons regarding how the statistics are compiled, their aggregated amount is distinguishable into banking-account- and trust-account-based figures. Aggregating the figures of the reporting banks[6] separately by trust banks and the rest of the banks, trust-account-based debt securities account for more than 90 percent of the external claims of the trust banks, while the share is almost 0 percent for the rest of the banks as of end-June 2018. These results suggest that almost all of the positions of the trust banks are held on a trust-account basis, whereas the positions of the rest of the banks are almost entirely on a banking-account basis (Chart 1). In what follows, the figures reported by trust banks are used as estimates of the positions held on a trust-account basis. Likewise, the figures reported by the rest of the reporting banks are treated as estimates of the banking-account-based positions in this paper.[7]

[Chart 1] Illustration of the figures of the reporting banks



Examining external claims and liabilities in the LBS with a breakdown of trustaccount- and banking-account-based figures defined above (Chart 2), the trustaccount basis accounts for more or less half of all external claims. Meanwhile, thebanking-account-based figures have large shares in external liabilities. This can be explained by the following characteristics of the reporting banks' activities over trust-account-based positions. First, many of them receive funds from domestic investors such as pension funds and trust funds in Japanese yen, where such positions are not included in the LBS since they are liabilities vis-à-vis residents denominated in Japanese yen. Second, such banks invest those funds in foreign assets, which are to be reported in the LBS since they are claims vis-à-vis non-residents. On a trust-account basis, the outstanding amount of external liabilities is therefore considered to be statistically small relative to those of external claims.[8]

[Chart 2] External claims and liabilities of Japanese banks (LBS)

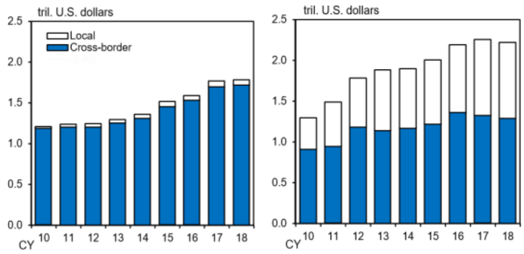


Note: The figures are adjusted for exchange rate fluctuations (to remove the impacts of foreign exchange rate changes, outstanding amounts denominated in major currencies other than U.S. dollar <Japanese yen, euro, British pound, and Swiss franc> are converted into U.S. dollar based on the exchange rates as of end-2017). Latest data as at end-June 2018.

Breaking down the foreign claims in the CBS into (1) cross-border claims and (2) local claims of overseas affiliates (Chart 3), the cross-border claims from entities located in Japan -- mainly debt securities -- account for almost all of the foreign claims held in a trust-account basis. On the other hand, on a banking-account basis, cross-border claims account for just less than 60 percent of foreign claims, reflecting increased credit activity by overseas affiliates.

[Chart 3] Foreign claims of Japanaese banks (CBS)

< Trust accounts > < Banking accounts>



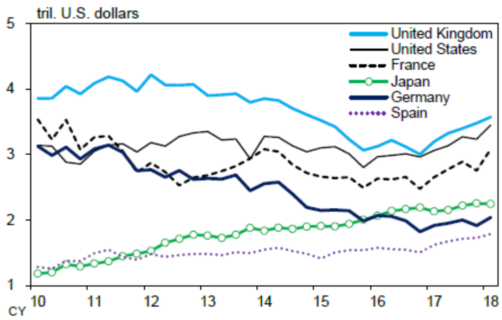
Notes: 1. The figures are end-of-year values, except for 2018 where end-June values are used.
2. Ultimate risk basis.
Source: BOJ.

3. International comparison using banking-account-based figures in Japan

This section provides an overview of international credit exposures held by Japanese banks. We use the estimated banking-account-based figures in Japan with the aforementioned procedure, and compare them with those of other major nationalities whose reporting figures are based on banking accounts.

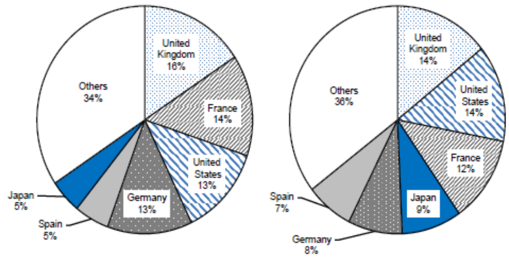
First, examining foreign claims in the CBS, the outstanding amount has recently been either more or less unchanged, or decreasing in many of the banks of other major nationalities. By contrast, that of Japanese banks has been rising substantially in recent years, standing at about 2.2 trillion U.S. dollars as of end-March 2018, marking the fourth largest share (9 percent) in the global market, following the United Kingdom, the United States, and France (Charts 4 and 5). Evaluating the ratio of foreign claims to the total assets

[Chart 4] Foreign claims held by banks of major nationalities (CBS)



Notes: 1. Latest data as at end-March 2018.
2. Ultimate risk basis.
Source: BOJ.

[Chart 5] Share of foreign claims by country of reporting banks (CBS)



Notes: 1. Latest data as at end-March values.
2. Ultimate risk basis.
Source: BOJ.

across countries (Chart 6), however, the ratio for Japanese banks is 17 percent as of end-March 2018, which is the lowest level among the banks of major nationalities in the banking sector. These result reveal that the ratio of external assets relative to the size of the banking sector is not significantly high on a global scale.

[Chart 6] Ratio of foreign claims held by banks of major nationalities (CBS)

	Foreign claims (bil. U.S. dollars) (A)	Total assests (bil. U.S. dollars) (B)	Ratio of foreign claims (A/B)
Japan	2,246.1	13,484.8	17%
United Kingdom	3,572.9	7,477.9	48%
Spain	1,787.2	4,182.6	43%
France	3,082.6	8,185.9	38%
Germany	2,039.9	8,651.8	24%
United States	3,450.7	15,057.2	23%

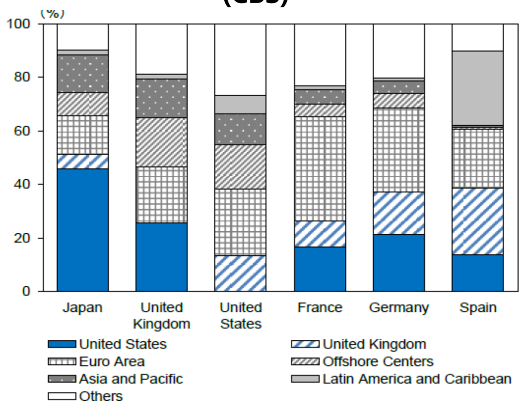
Notes: 1. The figures are as of end-March 2018.
2. Ultimate risk basis.
Source: BIS; BOJ.

Second, breaking down foreign claims by counterparty country (Chart 7), the Japanese banks' claims vis-à-vis the United States – such as positions in U.S. Treasuries and leading to local firms – account for more or less half of their overall foreign claims. For banks of other major nationalities, on the other hand, the shares of foreign claims are high on the following countries or regions: (1) the United States, major European countries, Hong Kong, and China held by U.K banks; (2) the United Kingdom, Japan, and the Cayman Islands held by U.S. banks; (3) the United States and the euro area held by French and German banks; and (4) "Latin America and Carribbean" held by Spanish banks. Next, in view of examining banks' claims on countries with which they do not share the same

domestic currency, we exclude claims vis-à-vis the euro area from the overall foreign claims held by French, German, and Spanish banks. The results show that, while the shares of claims vis-à-vis the United States held by French and German banks are 27 percent and 31 percent, respectively, the claims vis-à-vis the United Kingdom held by Spanish banks account for 32 percent of their foreign claims.

Third, determining an estimate of currency composition of consolidated-based foreign claims with the use of the LBS and the CBS[9] (Chart 8), U.S. dollar-denominated claims account for a large share – making up about 60 percent – of the foreign claims held by Japanese banks, reflecting the high share of claims held by French and German banks. Meanwhile, it show that the share of residual currencies – referred to as "Other" – is relatively high in the currency composition of Spanish banks. These currencies are considered to be the local currencies of "Latin America and Carribbean," which are the main counterparty countries of foreign claims held by Spanish banks. In this regard, from the point of view of examining claims denominated in non-local currencies, euro-denominated claims are now excluded form the total claims held by French, German and Spanish banks. Our findings are that the shares of U.S. dollar-denominated claims in their foreign claims are 54 percent, 63 percent, and 28 percent, respectively.

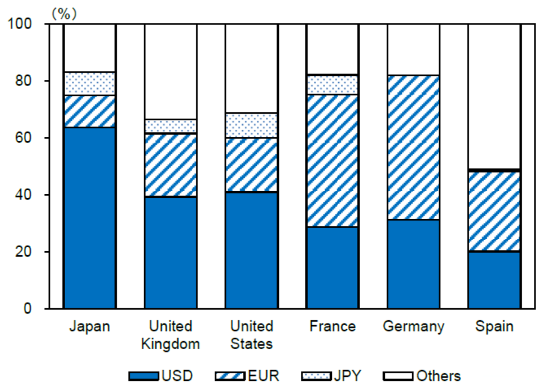
[Chart 7] Counterparty country and region composition of foreign claims (CBS)



Notes: 1. The figures are as of end-March 2018.
2. Ultimate risk basis.
3. "Asia and Pacific" and "Latin America and Caribbean" cover only developing countries located in corresponding regions.

Source: BIS; BOJ.

[Chart 8] Currency composition of foreign claims (LBS, CBS)

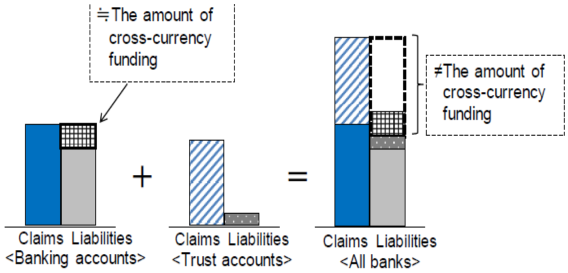


Notes: 1. The figures are as of end-March 2018.
2. JPY-denominated claims by German banks are included in "Others."
3. "Others" includes unallocated currency.

Source: BIS; BOJ.

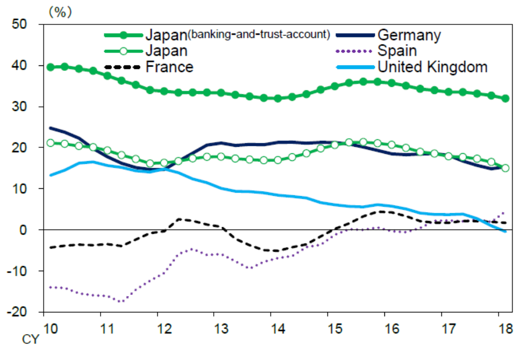
Fourth, we examine aspects of U.S. dollar funding. As an application of the IBS, the differential between U.S. dollar-denominated claims and liabilities is sometimes used as a proxy variable for the amount of cross-currency funding - funding by foreign exchange (FX) swaps and foreign currency swaps -- in the context of U.S. dollar funding. With regard to their positions on a trust-account basis, however, banks themselves do not use FX swaps or foreign currency swaps for foreign currency funding in many of such positions. The examples include positions in which institutional investors, such as pension funds, sell Japanese yen to buy foreign currencies, thereby investing them in foreign debt securities. Thus, there is a possibility that (1) taking the differential between the banking-and-trust-account-based international claims and liabilities held by Japanese banks, which include the aforementioned positions on a trust-account basis, and (2) using it as a proxy variable for the amount of cross-currency funding by Japanese banks, results in figures appearing to be greater than they actually are (Chart 9).[10]

[Chart 9] Relationship between (a) a differential between international claims and liabilities and (b) the amount of cross-currency funding



In this context, with the objective of examining cross-currency funding, we determine the ratios of (1) the differential between the U.S. dollar-denominated international claims and liabilities, relative to (2) the U.S. dollar-denominated international claims, for banks of major nationalities (Chart 10).[11] To derive the ratio for Japanese banks, banking account-based figures that we have estimated in this paper are used instead of banking-and-trust-account-based figures. The chart reveals that the banking-based ratio is significantly smaller than the banking-and-trust-account-based one for Japanese banks, indicating that the former is not extremely high on an international scale. Moreover, from a time-series perspective, their share has been more or less unchanged.[12]

[Chart 10] Ratio of the differential between the U.S. dollar-denominated international claims and liabilities relative to the U.S. dollar-denominated international claims



Notes: 4-quarter backward moving averages.
Latest data as at end-March 2018.
Sources: BIS; BOJ.

4. Conclusion Remarks

The IBS are fundamental statistics for capturing capital flows and credit exposures through internationally active banks. In the IBS, reporting practices of trustee business vary across reporting countries. While the BOJ reports figures that include positions both on trust-account and bankaccount bases, banks of other major nationalities do not include trust-account-based positions in their figures but they report figures based solely on banking accounts. Differences in the coverage of the positions should therefore be taken into account when carrying out international comparative analysis using the IBS. This paper has broken down the IBS in Japan into (1) positions held by trust banks, and (2) those of residual banks, and attempted to carry out an international comparison by treating the latter as an estimate of banking-account-based positions of Japanese banks as a whole.

Examining Japanese banks' international credit exposures with the use of the aforementioned estimates, while Japanese banks have been increasing their foreign claims in recent years, the ratio of their foreign claims relative to their total assets in the banking sector is not as high as those of banks of other major nationalities. It has also illustrated that Japanese banks' claims vis-à-vis the United States account for more or less half of their overall foreign claims. Likewise, in terms of currency, U.S. dollar-denominated claims account for a large share of the foreign claims held by Japanese banks. In the context of aspects of U.S. dollar funding, with the objective of examining cross-currency funding, we have determined the ratio of (1) the differential between the U.S. dollar-denominated international claims and liabilities, relative to (2) the U.S. dollar-denominated international claims, for banks of each major nationality using the IBS. The results have demonstrated that the banking-based ratio is significantly smaller than the banking-and-trust-account-based one for Japanese banks, indicating that the former is not extremely high on an international scale. Moreover, from a timeseries perspective, their share has remained more or less unchanged.

The BIS and the central banks worldwide have enhanced the IBS in the wake of the experience such as various global financial crises. The IBS enable one to provide a quantitative evaluation of past international capital flows and credit exposures through the banking sector with the breakdowns of their positions. The IBS are therefore considered to be growing in importance for analysis of Japanese banks, which have expanded the size of external claims, particularly in recent years. The BOJ will continue to contribute to further developments and enhancements of the statistics in cooperation with the BIS while paying concurrent attention to reporting burdens placed on reporting banks. Moreover, the BOJ will work on further enhancements to the usability and convenience of the IBS in Japan.

[1] The original paper is published as Saito, Hiyama and Shiotani (2018) "International Comparison of the BIS International Banking Statistics: Estimates of the Positions Excluding Trustee Business," Bank of Japan Review Series, 18-E-2

[2] For prior revisions of the IBS, please see the research papers and reviews below.
Bank of Japan Financial Markets Department (1999) "International Financial Markets as Viewed from BIS Statistics: Changes in the International Flow of Funds in the 1990s"
Nakahata, Kouda and Hishikawa (2002) "BIS Kokusai Yoshin Toukei no Tokuchouten to Minaoshi ni Muketa Torikumi" (Highlights of the BIS international banking statistics and Efforts for Enhancements) <Available only in Japanese>, Bank of Japan Market Review Series, 02-J-2
Hirano, Hayakawa, Saitou and Shigemi (2003) "Arata na BIS Kokusai Yoshin Toukei no Gaiyou--Yori Takakuteki na Country Risk no Haaku ni Muketa Kokusaiteki na Torikumi--" (Highlights of the New BIS international banking statistics -worldwide efforts for capturing country risks from various aspects-) <Available only in Japanese> Bank of Japan Market Review Series, 03-J-8
Inoue, Yoshizaki, Sasamoto and Shiotani (2017) "Enhancements to the BIS International Banking Statistics and Highlights of the Results of the Statistics in Japan," Bank of Japan Review Series, 17-E-1

[3] Please see here for the global results of the IBS with associated details.
(https://www.bis.org/statistics/about_banking_stats.htm)

[4]Please see here for the IBS in Japan with associated details.
(http://www.boj.or.jp/en/statistics/bis/ibs/index.htm/)

[5] With regard to the LBS and the CBS, reporting standards for the statistics are indicated in the reporting guidelines for BIS international banking statistics. While reporting of banking-and-trust-account-based figures is encouraged for the LBS, reporting practices vary across reporting countries.

[6] In principle, the reporting banks of the IBS in Japan are holders of Japan Offshore Market (JOM) accounts among the banks located in Japan.

[7] Since the trust-account-based figures estimated in this paper include banking-account-based positions of trust banks, there is an upward bias in the estimates of trust-account-based positions. In other words, there is a downward bias in those of the banking-account-based positions. These points should be considered for the analysis presented in this paper.

[8]Attention should therefore be given when carrying out international comparative analysis on the level of external claims or the differential between the external claims and liabilities. On the other hand, the impacts of their difference on the results are thought to be small in the case of a comparison among the levels of their external liabilities.

[9] It is possible to estimate the currency composition of consolidated-based foreign claims by merging the LBS and the CBS. For estimation procedures, please see the following paper.
McGuire, P. and G. von Peter (2009) "The US dollar shortage in global banking and the international policy response," BIS Working Papers, No. 291

[10] It should be noted that the banking-account-based differential of international claims and liabilities is not all funded by FX swaps and currency swaps. It rather broadly indicates the amount of cross-currency funding.

[11] In this estimation, the differentials between international claims and liabilities vis-à-vis each of banks and non-banks are used as a proxy variable. The ratios of (1) the differential between the U.S. dollar-denominated international claims and liabilities, relative to (2) the U.S. dollar-denominated international claims are then determined. For details, please see the following paper. IMF (2018) "A Bumpy Road Ahead," Global Financial Stability Report April 2018

[12] Please see "Financial System Report" for the evaluation on Japanese banks.
(http://www.boj.or.jp/en/research/brp/fsr/index.htm/

From-whom-to-whom matrices: Analysis of the sectoral interlinkages in Chile

Ivette Fernández, Alfredo Fuentes
Central Bank of Chile

Abstract
This paper presents the methodology followed by the Central Bank of Chile for measuring the sectoral interlinkages or from-whom-to-whom (FWTW) matrices of the following instruments for the period 2003-2018: Currency, deposits, debt securities, loans, money market fund shares and Non-MMF investment fund shares. In addition, the relationships between debtor and creditor sectors for financial instruments are shown. The FWTW matrices indicates which sectors are providing financing to which other sectors, by giving a breakdown of financing by holder and counterparty sector. The results expose, mainly, the important relationships of Deposit-taking corporations with Households and Non-financial corporations, for deposits, loans and debt securities.

Keywords
Flow of funds; sectoral accounts; financial accounts

1. Introduction

In Chile, the National Accounts by Institutional Sector (NA-IS) measures the current accounts, accumulation accounts and financial balance sheets[1], following the recommendations of the System of National Accounts 2008 (2008, SNA - FMI, 2008). These statistics have been compiled since 1996 at an annual frequency, with a lag of 15 months in their publication. As of 2011, data was measured at a quarterly frequency, with a lag of 115 days. At present, series have been available at a quarterly frequency since 2003 with a lag of 100 days.

In this context, the from-whom-to-whom matrices (FWTW) were published in 2013 for deposits and loans[2]. These data have been increasing in the last years, currently the FWTW matrices are published for currency, deposits, debt securities, loans, money market fund shares and Non-MMF investment fund shares. These data are published by flows (financial transactions) and stock (outstanding amounts).

---

[1] For more information, see: National Accounts by Institutional Sectors, NA-SI. Methodologies and Results 2005-2011q1" (CBCh, 2011).

[2] For more information, see: "Sectoral Interlinkages in Chile: An Implementation of National Accounts by Institutional Sector" (Fernández, 2013).

This paper focuses on disseminating the methodology and measurement of the FWTW and explaining the content and scope of this information. The presentation of the results is accompanied by a networking visualization tool, which allows the examination of interrelationships using charts that summarize the information contained in the matrices for outstanding amounts in the main FWTW matrices.

2. Methodology

In the measurement of NA-IS, the FWTW matrices or flows of funds are implicitly generated[3] and they are a tool in the balancing of the financial accounts. These matrices are a three-dimensional representation of the financial account defining the two parties to the transaction —the debtor and the creditor— and the financial instrument involved. This information allows to identify the financial relationships between institutional sectors, and thus it improves the understanding of the role of financial flows in the economy.

In general, information from different sources, especially financial statements, is transformed into concepts of national accounts and grouped by institutional sector and in the different transactions that make up the integrated framework of national accounts (current account, accumulation account, financial account, and financial balances sheets).

In particular, financial flows (financial account) are disaggregated into financial instruments, and in a large number of cases a counterparty of each transaction can be defined. From these items, it is possible to identify "who" is awarded a financial instrument, and "who" is providing this instrument, being the counterparty. For example, the mortgage loans granted by the banks correspond to information from the Banks sector, from the loans instrument and the counterparty is Households.

Once the financial accounts have been compiled by institutional sectors, an inter-sectoral balancing process begins with the aim of ensuring consistency throughout the system. In the balancing process, for each instrument an arbitration process is performed based on rules associated with the hierarchy of information sources, generating the FWTW matrices:

A. Currency: Considers the liabilities side of the Central Bank and the Rest of the World, and the data are reconciled with the financial information of the commercial banks and other sectors to measure the assets of the Households.
B. Deposits: Include liabilities of the Central Bank, commercial banks, and the Rest of the World; savings accounts are allocated as a household asset, while the remaining instruments in this category are determined

[3] In varied literature, the term flow of funds is interchangeable with financial account. In this paper these shall be differentiated, as in the System of National Accounts.

based on the information of the other sectors, the non-financial sector is considered the residual sector.

C. Debt securities: Consider the balance sheet information for determining the total issued by each sector, but the classification between short and long term is not very precise for either assets or liabilities. Therefore, as a support, the balancing process for this instrument is drawn from a security-by-security data base, an administrative record provided by the Central Securities Depository.
D. Loans: Consider the data from Commercial Bank and other sectors. The counterparty for longterm loans in foreign currency is assumed to be the Rest of the World. Mortgage and consumer loans are allocated to households and commercial loans to non-financial firms.
E. Money market fund shares[4]: Include the information by funds and according to their characteristics are distributed by counterparty. In this way, the information of Households (asset) and Rest of the World (asset) is defined.
F. Fund shares other than the money market[5]: Consider the characteristics of the funds to classify them by counterparty and define the data for Households and for the rest of the sectors.

Finally, the sources of information may vary depending on the institutional sector being measured. Mostly the information is made up of accounting statements of statistical units and supplementary data obtained from monitoring institutions. These sources are:

- Non-financial corporations: Information from the Financial Market Commission (CMF) and the Internal Revenue Service (SII).
- Deposit-taking corporations: Principally data from the CMF and financial statement of Central Bank of Chile.
- Other financial intermediaries and financial auxiliaries: Information from the CMF and the Pensions Supervisor (SP).
- Insurance corporations and Pension funds: Data from the CMF, the Superintendence of Health and the Pensions Supervisor.
- Money market funds: Data from the CMF.
- General government: Information from the General Comptroller Office, and the Superintendence of Social Security.

[4] This category includes the share of those investment funds that make their investments mainly in money market instruments and transferable debt instruments with a residual maturity of less than or equal to one year. This sector groups mutual funds type 1 and 2 of the classification assigned by the Superintendence of Securities and Insurance (SVS) corresponding to those with portfolios of less than or equal to 360 days.

[5] This category includes the shares of those investment funds in which the funds raised are invested mainly in financial assets lasting more than one year and in non-financial assets (real estate), so they are not substitutes of deposits. Includes mutual funds and investment funds.

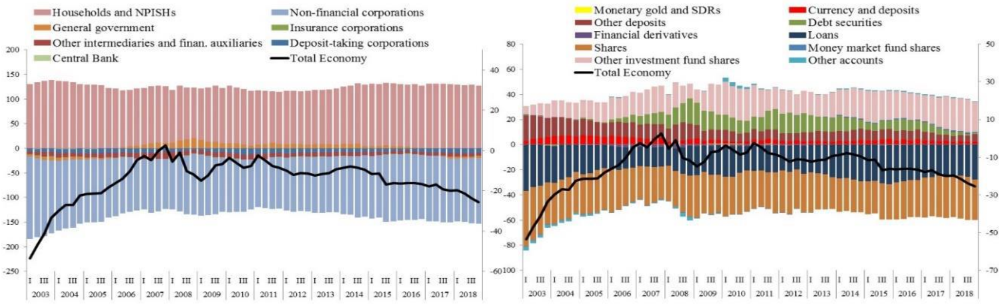
- Rest of the world: Data from the Balance of Payments and the International Investment Position from the Central Bank of Chile.

Additionally, other information sources are used directly from relevant agents in the securities market such as the Government or non-resident entities. In the last case, commercial banks that act as custodians of foreign investors report aggregate investments of these investors. Likewise, the securities custody information provided by the Central Securities Depository allows to reconcile the information of investment portfolios reported by pension funds, insurance companies, mutual funds and investment, among others.

3. Results

In Chile, the NA-IS provides an overview of the national economy by financial instruments and institutional sectors. The financial balance is influenced by the net lending position of the Households, in contrast with the rest of the sectors that present a net borrowing position (chart 1). Likewise, the balance sheet for financial instruments shows the importance of shares and other equity, and loans in net terms (chart 2).

Chart 1-2. Net Financial Assets by institutional sectors and financial instruments (percentage of annual GDP)

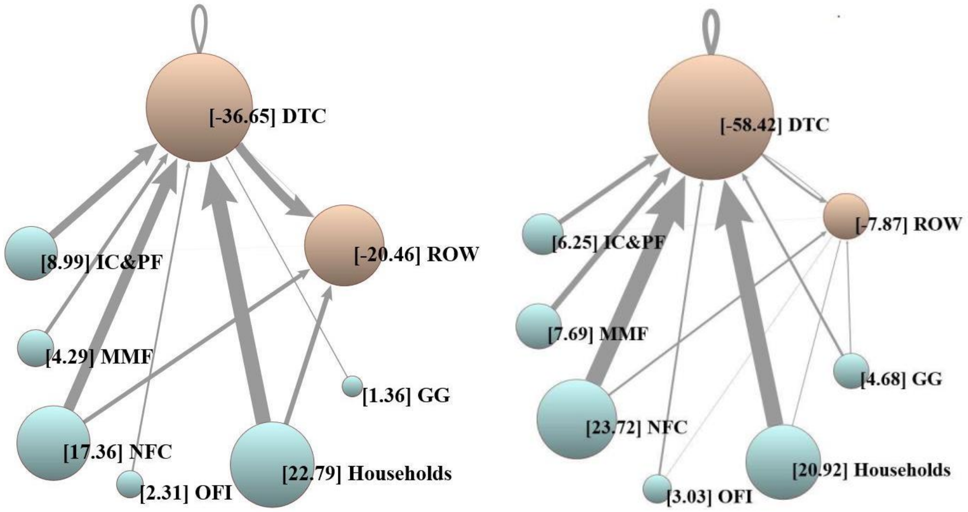


Source: Author's elaboration based on NA-IS data.

To know how the different sectors are related for some financial instruments, the "risk or network maps" are shown to visualize the relationships between these sectors and to compare results[6]. In deposits, Deposit-taking corporations increase the net stock and their interlinkages with other sectors between 2003 and 2018. Otherwise, the Rest of the world decreases both variables.

[6] The risk or network maps are drawn with Pajek software.

Chart 3. Sectoral interlinkages for Deposits. Outstanding amounts in December 2003 and 2018 (percentage of annual GDP)



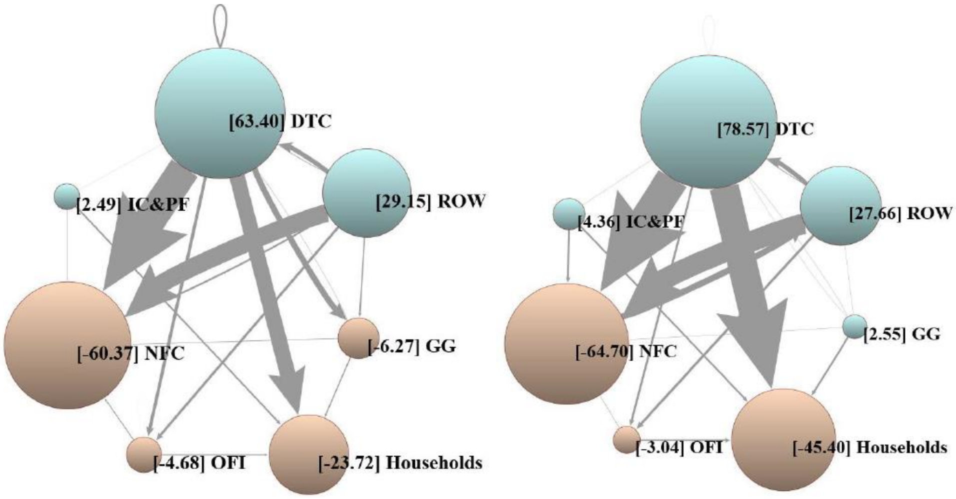
Source: Author's elaboration based on NA-IS data[7].

For loans, the outstanding amounts over the GDP increase both in stock and in their relationships for almost all sectors. Deposit-taking corporations with Households and Non-financial corporations have an important interrelationships, the banks finance both sectors. In the period, the Households increased 19.1 percentage points (pp.) and Non-financial corporations 4.1pp. (chart 4). On the other hand, the General Government decreased its stock of loans in the period of analysis with Banks and with the rest of the world.

[7] The magnitude of the lines indicates the intensity of the relationship; the arrows are pointing to the "depositories" or the direction of the money. The size of the circle indicates the net balance of deposits held by institutional sectors at the end of the period (deposits in assets minus deposits in liabilities). Orange circles indicate sectors with more liabilities than assets in deposits; and light blue circles indicate sectors with more assets than liabilities.

DTC: Deposit-taking corporations; **IC&PF**: Insurance corporations and Pension funds; **MMF**: Money market funds; **NFC**: Non-financial corporations; **OFI**: Other financial intermediaries; **Households**: Households and NPISHs; **GG**: General government; **ROW**: Rest of the world.

Chart 4. Sectoral interlinkages for Loans. Outstanding amounts in December 2003 and 2018 (percentage of annual GDP)



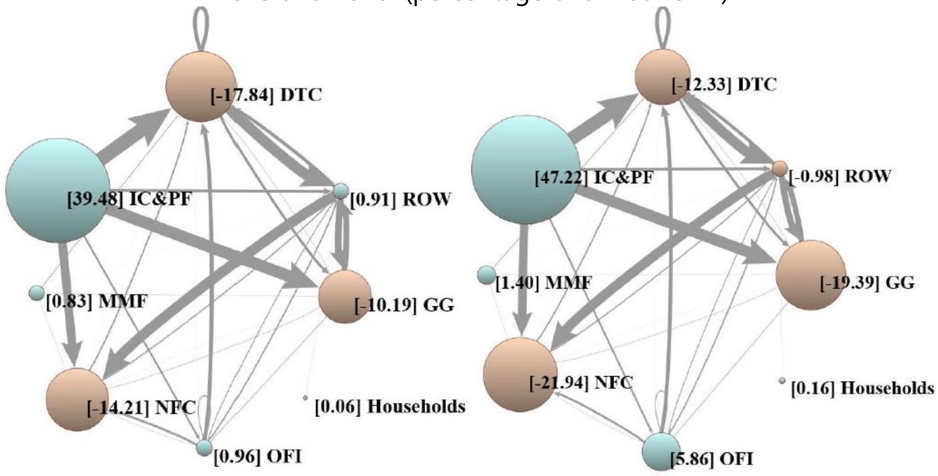
Source: Author's elaboration based on NA-IS data[8].

In the case of Debt securities, there are more interlinkages among sectors than other financial instruments. The ration over GDP has increased for some sectors as Insurance corporations and Pension funds during the analyzed period (chart 5).

[8] The magnitude of the lines indicates the intensity of the relationship; the arrows are pointing to the "borrower" or the direction of the money. The size of the circles indicates the loans maintained by the institutional sectors at the end of the period (loans granted minus loans received). Light blue circles indicate sectors with more assets than liabilities in loans and orange circles indicate sectors with more liabilities than assets of the same instrument.

DTC: Deposit-taking corporations; **IC&PF**: Insurance corporations and Pension funds; **NFC**: Non-financial corporations; **OFI**: Other financial intermediaries; **Households**: Households and NPISHs; **GG**: General government; **ROW**: Rest of the world.

Chart 5. Sectoral interlinkages for Debt securities. Outstanding amounts in December 2013 and 2018 (percentage of annual GDP)



Source: Author's elaboration based on NA-IS data[9].

4. Discussion and Conclusion

This paper presents the measurement methodology and the results of the FWTW matrices for currency, deposits, debt securities, loans, money market fund shares and Non-MMF investment fund shares

The FWTW matrix or flow of funds presents data that allows the analysis of financial instruments, which together with institutional sector statistics, provides information on sectoral interlinkages, indicating funding channels and systemic risks. Specifically, these matrices permit the study of structural changes in the balances of these instruments, and the cyclical changes in their flows over time, illustrating the relationships between different sectors.

One of the advantages of this information is that it is calculated within a unified and uniform system, this being the SNA 2008 in Chile. Furthermore, these matrices are comprehensive and include all sectors of the national economy and the relationships with the Rest of the world, allowing the analysis of both financing flows and their fluctuations over time.

---

[9] The magnitude of the lines indicates the intensity of the relationship; the arrows are pointing to the "bond issuers". The size of the circles indicates the debt securities maintained by the institutional sectors at the end of the period (asset minus liabilities). Light blue circles indicate sectors with more assets than liabilities of debt securities and orange circles indicate sectors with more liabilities than assets of the same instrument.

DTC: Deposit-taking corporations; **IC&PF**: Insurance corporations and Pension funds; **MMF**: Money market funds; **NFC**: Non-financial corporations; **OFI**: Other financial intermediaries; **Households**: Households and NPISHs; **GG**: General government; **ROW**: Rest of the world.

However, it is worth remembering that this information also exhibits some limitations in its use, due primarily to the fact that its publication does not cover all the financial instruments. In addition, to achieve an in-depth analysis it would be necessary to supplement the published data for these aggregated classifications of instruments and sectors with additional information; for example, through the complementary use of data from Household surveys which could facilitate an intra-sector analysis.

In conclusion, it will be necessary in the future to continue developing these matrices through the rest of the financial instruments, which will enable to represent the dynamics of savings and investment, investment and financing flows, and changes in investment portfolios between different sectors.

References

1. Central Bank of Chile (2011). "National Accounts by Institutional Sectors, NA-IS, methodology and results 2005-2011 q1 (only in Spanish)". Statistical Economic Studies, No. 87. Santiago, Chile.
2. Central Bank of Chile (2017ª). "Adoption of the SNA 2008 in the National Accounts by Institutional Sector (only in Spanish)". Methodological Note. Santiago, Chile.
3. Central Bank of Chile (2017b). "Benchmark Compilation 2013". Santiago, Chile.
4. Fernández, I. (2013). "Sectoral Interlinkages in Chile: An Implementation of National Accounts by Institutional Sector". Statistical Economic Studies, No. 97. Santiago, Chile.
5. Henríquez, C. – Pérez, J. (2016). "La Economía Chilena en el período 2003- 2014: Un Análisis desde la Perspectiva de las Cuentas Nacionales por Sector Institucional (only in Spanish)". Statistical Economic Studies, No. 117. Santiago, Chile.

The financial behaviour of the institutional sectors in the Euro area

Celestino Girón, Ilja Kristian KavoniusPierre Sola
European Central Bank[1]

Abstract
Over the first twenty years of the euro area, the member countries have shown very different dynamics in key financial variables. This has led analysts to group countries according to criteria that could allegedly help explain observed commonalities in such dynamics. However, the grouping criteria have not always been based on a sufficiently objective methodology, nor exploited the subtleness in country differences. We propose a grouping methodology based on an unsupervised machine learning clustering technology, Self-Organizing Map (SOM), combined with Principal Component Analysis (PCA) to obtain dimensionality reduction and interpretability of results. As a proof of concept, we apply the approach to classifying euro area countries on the basis of net lending by institutional sector. The results also enable a characterisation of the changes experienced in the distribution of imbalances by sector in the various countries as a result of the financial crisis in 2008.

Keywords
financial accounts; balance sheets; net lending/borrowing, sector imbalances; principal component analysis; deep learning; self-organising maps; clustering

1. Introduction

Over the last twenty years, analysts, researchers and the specialised media analysing economic developments in euro area countries have been grouping them on the basis of apparent similarities that allegedly had a bearing on the dynamics observed since the inception of the euro area, in particular after the financial crisis of 2008. Thus, countries have been grouped on the basis of relevant indicators, like indebtedness, leverage or housing price dynamics, or on more subjective criteria like north-south or core-periphery countries. In general, the grouping criteria have been ad-hoc and simple, and lacked the necessary objectivity in their choosing.

The country external position has been one of the criteria often used, typically distinguishing between countries that experience current account deficits and those with surplus. However, focusing on the total country net lending/borrowing ignores dramatic differences in the distribution of the

[1] The views expressed in this note are those of its authors and do not necessarily reflect those of the ECB.

imbalances across the economy. For instance, the same external deficit might be consistent with deficits in the public or in the private sectors and, in the latter case, with deficits for households or for corporations. Similarly, an external surplus might be accompanied with relative large deficits for some domestic sectors compensated by even larger surpluses for others. More in general, the distribution by sector and size of imbalances might present very different, complex configurations that might have a strong influence on the effects of shocks in the economy, and that make insufficient a country characterisation merely on the basis of the overall external deficit or surplus.

We propose a country characterisation and grouping criteria that take into account such differences. Given the potentially large complexity in the distribution of sector imbalances, any classification approach that relies only on human analysis would run the risk of suffering from “confirmation bias” or other cognitive biases. To partially correct this, we make use of a deep learning technique, Self-Organizing Map (SOM), a specific kind of neural network that identifies patterns in complex data.[2]

This approach can be applied to a much larger set of indicators, the analysis in this note being a “proof of concept” for such usage at a larger scale. However, interpreting the results in the way that is done across the note would not be possible with such larger indicator sets. To help bridge that difficulty we propose applying SOM to data previously transformed via Principal Component Analysis (PCA) which allows reducing the dimensionality of the outcome (although at the expense of decreasing the direct interpretability of the components/ indicators examined) . Even though we work here with a reduced set of indicators, we also apply PCA to it in a second exercise to provide a geometric interpretation of the SOM results.
The next section of this paper discusses briefly the applied methodology. The third section shows the data analysis. The last section summarises and briefly discusses avenues for future work.

2. Methodology

Our data set consists of quarterly time series for the net lending/net borrowing of the individual domestic sectors - non-financial corporations (NFCs), financial corporations, general government and households (including non-profit institutions serving households)- for all countries in the euro area. We work with medians of four quarter rolling sums expressed in percentage of GDP. The medians are calculated for the ranges1999-2008 and 2009-2018, i.e. for the pre-crisis and crisis and post-crisis period, when the distribution of sector imbalances showed large changes. We then have 38 countryperiod

[2] Similar approaches have been applied by: López Iturriaga and Pastor 2013. Peltonen and Sarlin 2011.

pairs (two different periods for 19 countries) represented by 4 indicators corresponding to the sector net lending medians. This organisation of the information allows not only to cluster countries according to similarities in the distribution of imbalances, but also to examine transitions from one cluster to another as a result of the financial crisis.

We cluster these 38 country-period pairs using SOM, which yields a simplified representation of the input space of 4-dimensional net lending vectors corresponding to the pairs. The input vectors are mapped onto a network of a few, also 4-dimensional output vectors ("neuron" weights vectors) whose topological configuration is as close as possible to that of the input space (the network is "trained" on the basis of the input data). SOM thus creates a low-resolution view of the larger data set, which, apart from yielding a clustering on the basis of similarities in the input data, is useful for the purposes of visualization and general data analysis.[3] In our set-up we have used a 2x3 hexagonal neural network to map the input space. We believe that the resulting clustering into six neurons/categories strikes a balance between delivering a manageable number of clusters and still allowing for detecting "outliers", country-period pairs that do not present clear analogies with others.

As a second step in our analysis, we use additionally PCA, which is applied to the initial data to obtain a second data set to which SOM is then applied. PCA transforms a set of observations of possibly correlated variables (in our case net lending per sector) into other variables called "components". This transformation is defined in such a way that the components get ordered by variance, the first "principal" component having the largest variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn having the highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting components (each being a linear combination of the original variables) are then an uncorrelated orthogonal basis set.[4]

PCA is a useful tool for identifying key characteristics of a large set of indicators, based on their variance, and synthetizing them into a reduced number of "components" (those with the highest variance) thus reducing dimensionality and helping in analysing and visualising the information. In this paper we use PCA for obtaining results that can be depicted in a two-dimensional chart (see below).

3. Results

Figure 1 shows the outcome of the clustering exercise on the original data. The country-period pairs are denoted with the ISO 3166-1 alpha-2 country

[3] Kohonen 1982, 59–69.

[4] Pearson 1901, pp. 559-572.

code and the suffixes 1 or 2 for the two consecutive periods. They are shown for each cluster together with two 4xn matrices that contain the *standardized*[5] 4-dimensional vectors with the medians of sector net lending/ net borrowing for n countries included in the cluster for each of the two time periods.[6] For each cluster, the corresponding "neuron weight vector" is provided, i.e. the closest node in the "trained" neural network, which can be interpreted as the (standardised) net lending/ net borrowing configuration of a notional representative country for that cluster.

Figure 1. SOM country clustering on the basis of domestic sector net lending/ net borrowing

Cluster	Neuron	Countries
1	$\begin{bmatrix} S11 \\ S12 \\ S13 \\ S1M \end{bmatrix} = \begin{bmatrix} 0.61 \\ 0.01 \\ 1.19 \\ -0.51 \end{bmatrix}$	1999-2008: **BE_1, FI_1, IE_1, LU_1, NL_1** $\begin{bmatrix} 0.23 & 1.62 & 0.29 & 0.32 & 1.75 \\ 0.10 & -0.29 & 0.62 & -0.28 & 0.22 \\ 0.85 & 2.43 & 1.42 & 1.67 & 0.73 \\ 0.50 & -0.74 & -1.96 & -0.44 & -0.66 \end{bmatrix}$ 2009-2018: **EE_2, FI_2, DE_2, LU_2** $\begin{bmatrix} 0.12 & 1.08 & 0.86 & -0.06 \\ -0.15 & -0.27 & -0.42 & 0.09 \\ 0.89 & 0.14 & 0.98 & 1.39 \\ 0.29 & -0.82 & 0.72 & -024 \end{bmatrix}$
2	$\begin{bmatrix} 0.32 \\ 0.08 \\ -0.49 \\ 0.24 \end{bmatrix}$	1999-2008: **AT_1, FR_1, DE_1, IT_1** $\begin{bmatrix} 0.11 & 0.31 & 0.36 & 0.06 \\ -0.26 & 0.01 & -0.11 & 0.07 \\ 0.14 & -0.13 & -0.40 & -0.38 \\ 0.61 & 0.21 & 0.81 & 0.18 \end{bmatrix}$ 2009-2018: **AT_2**, **BE_2, FR_2, IE_2, IT_ 2, LV_2, LT_2, NL_2, PT_2, SK_2, SI_2, ES_2** $\begin{bmatrix} 0.32 & 0.31 & 0.20 & 0.31 & 0.32 & 0.34 & 0.54 & 1.47 & 0.12 & 0.22 & 0.56 & 0.88 \\ 0.16 & 0.20 & -0.07 & 0.65 & 0.31 & -0.33 & 0.01 & 0.32 & 0.50 & -0.31 & 0.15 & 0.29 \\ 0.15 & -0.29 & -0.61 & -1.13 & -0.21 & 0.34 & 0.12 & 0.09 & -1.21 & -0.15 & -0.90 & -1.40 \\ 0.18 & -0.04 & 0.26 & 0.03 & -0.06 & 0.34 & 0.02 & 0.20 & 0.12 & 0.01 & 0.60 & 0.10 \end{bmatrix}$
3	$\begin{bmatrix} -1.26 \\ 0.46 \\ -0.62 \\ 0.50 \end{bmatrix}$	1999-2008: **MT_1, PT_1, SK_1, SI_1** $\begin{bmatrix} -1.63 & -1.00 & -1.17 & -1.15 \\ 2.54 & -0.05 & -1.14 & -0.06 \\ -1.03 & -0.79 & -0.98 & 0.01 \\ 0.61 & 0.01 & 0.66 & 0.61 \end{bmatrix}$ 2009-2018: **CY_2**, **MT_2** $\begin{bmatrix} -0.66 & -1.97 \\ 1.30 & 0.34 \\ -0.76 & -0.14 \\ -0.51 & 1.58 \end{bmatrix}$

[5] i.e. expressed as differences from the mean divided by the standard deviation.

[6] Sectors codes are those of the SDMX standard, i.e. S11 NFCs, S12 financial corporations, S13 general government and S1M households and non-profit institutions serving households.

4	$\begin{bmatrix}-1.60\\0.02\\0.78\\-0.92\end{bmatrix}$	1999-2008: **EE_1, LV_1, LT_1, ES_1**	$\begin{bmatrix}-1.80 & -2.79 & -1.21 & -0.58\\-0.08 & 0.16 & 0.05 & -0.04\\1.48 & 0.47 & 0.38 & 0.78\\-1.16 & -1.38 & -0.75 & -0.39\end{bmatrix}$
		2009-2018: *None*	
5	$\begin{bmatrix}1.07\\0.39\\0.78\\-0.92\end{bmatrix}$	1999-2008: **GR_1**	$\begin{bmatrix}0.86\\0.21\\-1.47\\-1.81\end{bmatrix}$
		2009-2018: **GR_2**	$\begin{bmatrix}1.28\\0.57\\-2.50\\-1.54\end{bmatrix}$
6	$\begin{bmatrix}-0.39\\-4.98\\0.02\\3.89\end{bmatrix}$	1999-2008: **CY_1**	$\begin{bmatrix}-0.39\\-4.98\\0.02\\3.89\end{bmatrix}$
		2009-2018: *None*	

The neuron in cluster 1 ($\begin{bmatrix}0.61\\0.01\\1.19\\0.51\end{bmatrix}$) represents countries with high NFC net lending (0.61; it may be recalled that values are standardised, i.e. expressed as differences from the mean divided by the standard deviation), average financial sector net lending (0.01), large government surplus (1.19) and households net lending/ borrowing below average (-0.51). The countries clustered here broadly follow that pattern (BE, FI, IE, LU, NL before the crisis), although not exactly, presenting variations in particular in the financial sector. In any case, the methodology followed ensures that the countries grouped under the clusters are closer to the corresponding neuron than to any other neuron. Note that an increase in the number of clusters would deliver lower differences between countries and neurons, but also note that too many clusters would result in an impractical outcome for classification purposes (in the limit, having as many clusters as countries would deliver perfect neuron-country matches, but this would not be of any use). As stated above, our choice for six clusters tries to balance out the need to have sufficiently low resolution in the clustering and at the same time to avoid grouping countries that are too different from each other.

Cluster 2 ($\begin{bmatrix}0.32\\0.08\\-0.49\\0.24\end{bmatrix}$) encompasses countries with relatively high NFC lending (0.32), although not as high as in cluster 1, and average financial sector lending, just as cluster 1. However, the neuron shows a dramatic difference for the fiscal situation compared with cluster 1, with government deficit above the average (-0.49) as opposed to below the average (or to surplus). At the same time, households present relative high net lending (0.24), rather than net borrowing as in cluster 1.

Cluster 3 ($\begin{bmatrix}-1.26\\0.49\\-0.62\\0.50\end{bmatrix}$) represents a textbook net lending configuration, with deficits in non-financial corporations (-1.26) and government (-0.62) and surpluses in the financial sector (0.49) and households (0.50), although the former are larger than the latter, leading to an overall external deficit. The neuron in cluster 4 ($\begin{bmatrix}-1.60\\0.02\\0.78\\-0.92\end{bmatrix}$) is a more extreme deficit economy stemming from the private sector (-1.60 for NFCs, -0.92 for households), while the government presents higher net lending (0.78).

Clusters 5 and 6 represent outliers, GR before and after the crisis (neuron $\begin{bmatrix}1.07\\0.39\\-1.98\\-1.68\end{bmatrix}$ with contemporaneous deficits in government and households not seen in any other country, and CY before the crisis ($\begin{bmatrix}-0.39\\-4.98\\0.02\\3.89\end{bmatrix}$), due to the exceptionally high deficit in the financial sector (-4.98 standardised).

The clustering presents borderline cases, which can be detected by training the SOM a few times and identifying cases where the pair country-period is sometimes allocated to different clusters (SOM, as other deep learning technologies, does not deliver the exact same final results every time they are run, these being dependant on the initialization of the neural network[7]). One prominent case is ES before the crisis (ES_1), which shows deficits in the private sector of significantly lower magnitude than those of the other countries in the cluster (EE, LV, LT). Some other clustering results put ES_1 in cluster 1, rather than in cluster 4, on the basis of its relatively better fiscal position (in such a case the neuron of cluster 4 also becomes more tilted towards the ES_1 configuration with lower NFC and government relative surpluses).

Now we move to the second part of the analysis and apply the PCA before the SOM. As mentioned earlier, this allows us to work with reduced dimensionality and better visualise the closeness of the country-period pairs

---

[7] This can be avoided by setting fix weights for the initial neural network before training. However we have followed random initialization precisely to obtain information on possible borderline cases.

to the neurons of the SOM, so that, for instance, we can appreciate borderline cases like ES_1. Figure 2 shows both the trained neuron network (in red, with neurons labelled with the cluster numbering in Figure 1) and country-period values (or "scores", in green) for the first two components of the PCA transformation of the original data, covering around 67% of the total variability in the data set. Selected country-period pairs are also marked.

Figure 2. First two components of PCA representation of sector net lending/net borrowing. SOM neuron weights and country-period scores

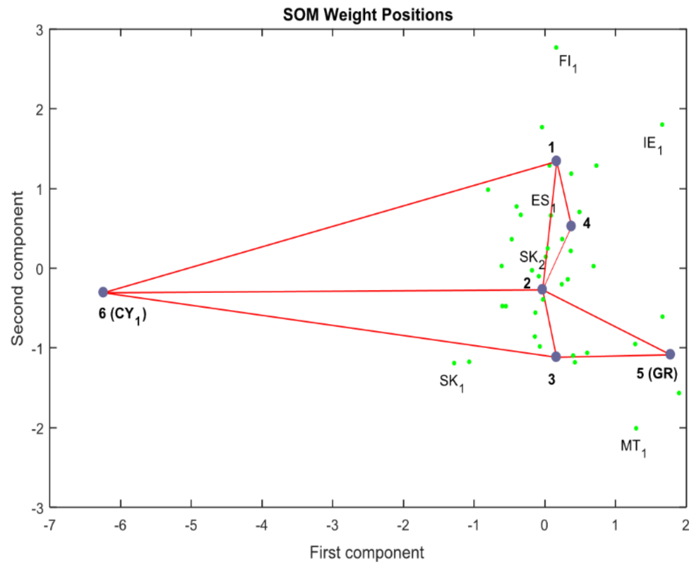


Figure 2 allows a meaningful 2-dimensional representation and interpretation of the clustering outcome that would not be possible with the original data (although this is an incomplete representation as it does not cover the remaining 33% of sample variance captured by the other two components). The outliers, in particular CY_1, can be clearly identified as showing scores far away from all the others. At the same time, we can observe a relative higher disparity within some clusters. Cluster 1 contains pairs as far away from the central neuron as FI_1 and IE_1, which are also very far away from each other mainly due to differences in the net lending of the financial sector and households. Similarly, MT_1 and SK_1 in cluster 3 show very different imbalances for the financial sector, visualised as a relative high distance between them.

One may also notice in Figure 2 the transition made by SK from cluster 3 before the crisis corresponding to countries with deficits in government and NFCs- to cluster 2 after the crisis. Figure 3 provides complete information on such transitions, with country codes in white representing the country before the crisis, and in yellow the country after the crisis in case such country has gone through a cluster transition. The Figure includes a broad characterisation of the clusters on the basis of the net lending of NFCs (S11) and government (S13) to help follow the changes undergone in domestic imbalances by the countries in transition.

Figure 3. Cluster transitions after the crisis *(positions of countries within the coloured areas do not convey information on "closeness" between them in the input space)*

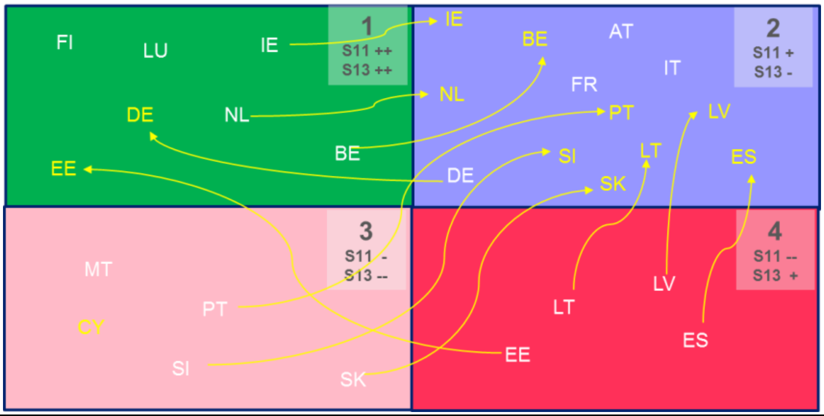


Figure 3 shows that almost all countries converge like SK to cluster 2, characterised by government deficit and NFC surplus, in particular those that experienced NFC deficits before the crisis and were included in cluster 4, which becomes empty, and cluster 3, which remains only with MT and CY (moved in from the outlier cluster 6). An exception to this apparent attraction force exerted by cluster 2 on countries with NFC deficits is EE, which moves to cluster 1 with even stronger NFC net lending and yet government surplus. A very relevant transition is that of DE, also to cluster 1 from the now increasingly "normal" cluster 2.

4. Discussion and Conclusion

In this paper we have explored how an objective quantitative approach can be used to cluster countries according to their net lending/ net borrowing configurations, and to show how the financial crisis starting in 2008 affected that clustering. In particular we have used SOM, a deep learning technology that allows, inter alia, extracting patterns from complex data.

The results of the clustering and their changes in time make sense from an economic perspective. FI and LU are continuously grouped together as countries that fundamentally present large NFCs surpluses and good fiscal positions; they were joined after 2008 by DE and EE as a response to the crisis shock. AT, FR and IT are continuously clustered in a group with still low NFC deficits, but relative large government deficits, situation that was shared by almost all countries after the crisis (BE, ES, IE, LT, LV, NL, PT, SK and SI) most of them previously clustered as countries experiencing booms.

This exercise has been run with a limited number of indicators, the sector imbalances, as a proof of concept for the proposed approach based on SOM. Future work will extend the methodology to a larger number of indicators in order to cover different aspects of the financial situation, e.g. indebtedness, assets and net worth, as well as holding gain and losses, all in order to deepen the analysis of financial developments. Such extensions will increase remarkably the dimensionality of the exercise. To help in the analysis of the results, the data will be pre-processed through PCA. This note also contains an application of this extension to improve the visualisation of the results of the clustering on the basis of sectoral net lending.

References

1. *Kohonen, T. (1982): "Self-Organized Formation of Topologically Correct Feature Maps". Biological Cybernetics. 43 (1): 59–69.*
2. *López Iturriaga, F. J. and I. Pastor (2013): "Self-organizing maps as a tool to compare financial macroeconomic imbalances: The European, Spanish and German case". Spanish Review of Financial Economics. Vol. 11 (2): 47-94.*
3. *Pearson, K. (1901): "On Lines and Planes of Closest Fit to Systems of Points in Space". Philosophical Magazine. 2 (11): 559–572.*
4. *Peltonen, T. and P.; Sarlin (2011): "Mapping the state of financial stability". European Central Bank Working Paper series no 1382. Available at* https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1382.pdf (accessed April 2019).

On the discriminatory and predictive accuracy of RDT against the microscopy smear tests in the diagnosis of malaria among under-five children in Nigeria

Fagbamigbe, A. F.

Abstract

Accurate identification of malaria cases is very crucial to the management of cases and to the success of malaria eradication agenda. This study is designed to evaluate the discriminatory and predictive accuracy of Malaria Rapid Diagnostic Tests (RDTs) in Nigeria. The data obtained during the 2015 Nigeria Malaria Indicator Survey was used to quantify the discriminatory accuracy through the analysis of its sensitivity, specificity, positive (LR+) and negative (LR-) likelihood ratio. The positive (PPV), and negative (NPV) predictive values, area under the receiver operating characteristic curve, and diagnostic odds ratio were used to assess the predictive accuracy of the RDTs using expert microscopy as a gold standard at p=0.05. The McNemar paired test and the Kappa statistics were used to assess agreement between the diagnostic tests. There was a significant but not an excellent agreement between the results of the RDT and microscopy tests (p<0.001). The overall sensitivity of the RDT was 87.6%(85.9%-89.2%), specificity was 75.8%(74.4-77.1%) while the diagnostic accuracy stood at 79.0%(77.9%-80.0%). The LR+, LR-, PPV and NPV were 3.6(3.4-3.8) and 0.16(0.14-0.19), 57.5%(56.1%-58.9%) and 94.2%(93.5%-94.9%) respectively. The sensitivity of RDT increased as the age of the children increased, from 85.7% among those aged 0-6 months to 86.1% in 7-23 month olds to 88.1% among those aged 24-59 months but the reverse was the specificity. For children with severe anaemia, the sensitivity of the RDT was nearly 100% compared with a specificity of 39%. While the sensitivity and the PPV reduced with children's level of anaemia, the higher the severity of anaemia the lower the NPV, specificity, the diagnostic accuracy of the RDT. The odds of RDT been sensitive was about 50% (aOR=0.52(95% CI: 0.30-0.90)) lower among children aged 7-23 months compared with those aged 24-59 months while the odds of RDT been sensitive was 2 times (aOR=2.15 (95% CI:1.67-2.77)) higher among those 7-23 months than among those aged 24-59 months. Although there was a significant agreement in the outcomes of RDT and Microscopy tests, the discriminatory accuracy of RDT was weak. Also, the predictive accuracy, especially the PPV of the RDTS, were very low. These measures of accuracies differed across the age of the children, level of anaemia, and other characteristics. Without an accurate, efficient and reliable diagnosis of malaria, the goal of eliminating malaria and reduction of malaria-related deaths to zero by 2020 will only be a wide chase.

Keywords
Rapid Diagnostic Tests; Microscopy; diagnostic accuracy; sensitivity; specificity

1. Introduction

Malaria remains a major public health challenge in sub-Sahara Africa, especially Nigeria [1–3]. Despite myriad efforts devoted to curbing the menace of malaria in Nigeria, its prevalence has remained high [4] with nearly half of Nigeria 20 million under-five children infected by malaria parasite on daily basis. Some of the malaria eradication initiatives taken by Nigeria government and other stakeholders include mass long-lasting insecticidal net (LLIN) campaigns and distribution, replacement campaigns, intermittent preventive treatment (IPT), and a massive scale-up in malaria case management including use of RDTs.

Malaria diagnosis is a key pillar in the eradication of malaria in Africa. WHO recommends that malaria case management is based on parasite diagnosis in all cases [7,13] and that treatment should only commence after diagnosis [14]. The use of antigen detecting Rapid Diagnostic Tests (RDTs) is a vital part of this strategy. Malaria RDT is the "immunochromatographic lateral flow device for the detection of malaria parasite antigens [7,15,16]. The strategic purpose for the introduction of RDT is to extend access to malaria diagnosis by providing a parasite-based diagnosis in areas where "good-quality" microscopy cannot be maintained [7], unavailable or less convenient; hence case management is improved [11]. To enhance effective diagnosis of all malaria cases, the diagnostic method used must be accurate and available at the point of care.

Literature has documented the specificities, sensitivities, numbers of false positives and false negatives and variabilities in temperature tolerances of these tests as some of the difficulties and challenges facing current RDTs [16]. Conclusive evidence is still lacking on the accuracy and safety of a testbased strategy for children [14]. It is, therefore, imperative to evaluate the performances of RDTs in terms of diagnostic accuracy as this measure provides information on the diagnostic test's ability "discriminate between and/or predict disease and health" [18].

The consequences of erroneous and wrong diagnosis are enormous. These errors could be either false negative or false positive errors. The false negative errors occur when the disease is missed when indeed, it is present. They often result in people foregoing needed treatment and could lead to the chronic stage of disease or even death. The false positive errors are due to wrongful confirmation that a disease is present. It leads to a wrong focus on disease-free subjects which may result in unnecessary treatment and sometimes overtreatment. It could lead to a negative impact, personal inconvenience, unwarranted stress, anxiety etc. Eusebi et al have already advocated for strict evaluation of the diagnostic accuracy of testing procedure aimed at validating any potential diagnostic tool [18]. This study is, therefore, designed to provide the malaria programmers with situation assessment of the accuracy of RDTs being used in Nigeria and to also determine the distribution of the levels of accuracy in terms of discrimination and prediction. It was hypothesized in this study that the outcomes of both the RDT and microscopy malaria tests are not in agreement. The findings in this study will improve programme implementation and enhance the much-needed progress towards malaria control and eradication in Nigeria.

2. Methods

Malaria testing using RDT

Using the same blood sample collected for anaemia testing, a drop of blood was tested immediately with the SD BIOLINE Malaria Ag P.f (HRP-II)™ (Standard Diagnostics, Inc.) RDT, being a qualitative test “to detect histidine-rich protein II antigen of Plasmodium falciparum (Pf) in human whole blood” [4]. The test procedures were handled by well-trained field laboratory scientists in accordance with the RDTs’ manufacturer’s instructions. The RDT results were provided to each child’s parent or guardian in oral and written forms within 15 minutes and were recorded on the Biomarker Questionnaire. Children that tested positive to malaria and not currently on treatment with artemisinin-based combination therapy (ACT) or who had not completed a full course of the ACT during the preceding 2 weeks were given full treatment according to the Nigeria national malaria treatment guidelines[4].

Malaria testing using blood smears

In addition to the RDT, thick and thin blood smears were prepared in the field. Each blood smear slide was labelled according to guidelines and transmitted to the laboratory. The thick and thin smear slides were stained at zonal staining and taken to the ANDI Centre of Excellence for Malaria Diagnosis, University of Lagos, Nigeria for logging and microscopic reading. Other details of the testing procedures have been reported earlier[4].

Description of variables

The outcome variable in this study is the result of the RDT and microscopy malaria tests while the independent factors considered are child’s household wealth quintiles, child age, and sex of children, mother’s educational attainment, place of residence, region, sleeping under a long-lasting insecticide-treated net or any ever treated nets recently, experience of fever within 2 weeks preceding the survey, and the level of anaemia as used in earlier studies[3,20]. The ages of the children were categorized into 0-6, 7-23, and 24-59 months as used in earlier studies on under-five children [21,22].

3. Data Analysis

There were a total of 7,011 children aged 6-59 months across all the households visited during the survey. Basic descriptive statistics were used to describe the under-five children with respect to the characteristics of their mothers. The McNemar paired test and the Kappa statistics were used to test the hypothesis of non-agreement and to determine the level of agreement between the outcomes of the diagnostic tests respectively. Diagnostic test parameters including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) was used to determine the accuracy of the malaria RDT results in comparison with the "Gold standard" results from microscopy tests. Let X denote the true state of a person (microscopy test results) with microscopy positive = D+ and negative = D-. Also, let Y be the outcome of the RDT test, with RDT positive = T+ and negative = T-. Then,

Sensitivity = $P(Y = \mathrm{T}^{+}|\mathrm{X} = \mathrm{D}^{+})$; Specificity = $P(Y = \mathrm{T}^{-}|X = \mathrm{D}^{-})$. The positive likelihood ratio (LR+) and the negative likelihood ratio (LR-) are calculated as

$$\mathrm{LR+}=\frac{TPR}{FPR}=\frac{Sensitivity}{1-Specificity} \text{ and } \mathrm{LR-}=\frac{FNR}{TNR}=\frac{1-Sensitivity}{Specificity}$$

as proposed by Simel et al. [23]. The unique statistics produced by the likelihood ratios have made it the optimal choice for reporting diagnostic accuracy for clinically meaningful thresholds[18].

However, there is a need to determine the predictive values since sensitivities and specificities are not measures of prediction[18]. Predictive values depend on disease prevalence, and their conclusions are transposable to other settings. The predictive values help to determine how likely the disease is, given the test result. The PPV is the probability that the disease is present, given that the diagnostic test is positive. It is computed as TP/ (TP+FP) while the NPV is the probability that the disease is not present given that the test is negative, computed as TN/ (TN+FN). A diagnostic test could be said to be perfect if it can predict perfectly, i.e., if PPV = NPV=1. The PPV decreases with decreasing prevalence.

$$\mathrm{PPV} = P(\mathrm{D}^{+}|\mathrm{T}^{+});\ \mathrm{NPV} = P(\mathrm{D}^{-}|\mathrm{T}^{-}).$$

The accuracy of a test = (TP+TN)/(TP+TN+FP+FN) and its confidence intervals are the standard logits as given by Mercaldo et al[24]. The pretest probability is the same as the prevalence as determined by the gold standard, the pretest odds is prevalence/(1-prevalence), posttest odds= pretest odds *likelihood ratio and the posttest probability=posttest odds/(1+positive odds).

The ROC analysis is used in diagnostic screening evaluation to quantify the accuracy of diagnostic
95 tests [25].

Results

Among the 7,011 children included in the survey, only 6,025 and 5,753 children were tested for malaria parasite using the microscopy and the RDT methods respectively. The distribution of the results of the tests by selected children characteristics is presented in Table 1. For the two test procedures, malaria prevalence increased with the children ages and was insignificantly higher among male children than among female children. In all, children residing in the rural areas, Northern geopolitical zones, who came from homes in the poorer wealth quintiles and who slept under treated nets recently and who had fever within the two weeks preceding the survey had a higher significant prevalence of malaria for both the RDT and microscopy malaria tests.

In addition, the level of anaemia was significantly associated with the presence of malaria parasite as evidenced by both the RDT and microscopic tests. Children with severe anaemia had higher malaria prevalence (RDT+= 88%; microscopy+=67%) compared with others that were not anaemic (RDT+ =25%; microscopy+=13%). Although not stated in the tables, *Plasmodium falciparum* was the main (94%) type of malaria parasite found through the blood film microscopy among the U5 children, followed by the *Plasmodium ovale* and *Plasmodium malaria* at 5% and 1% respectively.

Table 1: Distribution of the results of the malaria tests using RDT and microscopy and their association with selected children characteristics in the Nigeria 2015 MIS

Children Characteristic	Number (n=7011)	%	Malaria Test Results			
			RDT+ (%)	x^2 p-value	Micro+ (%)	x^2 p-value
Child's age (months)						
0-6	755	10.8	21.6	*<0.001	10.0	*<0.001
7-23	1 914	27.3	36.2		20.3	
24-59	4 342	61.9	49.5		30.9	
Sex						
Male	3 574	51.0	46.2	0.063	27.8	0.276
Female	3 437	49.0	44.0		26.9	
Location						
Urban	2 350	33.5	24.1	*<0.001	11.4	*<0.001
Rural	4 661	66.5	55.7		35.6	
Zone						
North Central	1 309	18.7	50.7	*<0.001	32.1	*<0.001
North East	983	14.0	42.9		26.5	
North West	2 286	32.6	58.2		37.1	
South East	601	8.6	31.7		13.9	

Children Characteristic	Number (n=7011)	%	Malaria Test Results			
			RDT+ (%)	x^2 p-value	Micro+ (%)	x^2 p-value
South South	780	11.1	28.9		19.4	
South West	1 053	15.0	32.1		15.3	
Usual resident						
No	69	1.0	36.8	0.147	25.2	0.598
Yes	6 942	99.0	45.2		27.4	
Wealth Quintile						
Poorest	1 474	21.0	64.1	*<0.001	43.1	*<0.001
Poorer	1 612	23.0	62.7		41.0	
Middle	1 333	19.0	49.2		27.7	
Richer	1 289	18.4	30.2		16.8	
Richest	1 303	18.6	12.7		4.3	
Slept under ever treated net						
No	3 963	56.5	42.1	*<0.001	25.7	*<0.001
Yes	3 048	43.5	49.0		29.5	
Slept under LLITNs						
No	3 980	56.8	42.0	*<0.001	25.7	*<0.001
Yes	3 031	43.2	49.1		29.5	
Fever in last 2 weeks						
No	4 123	59.0	37.4	*<0.001	23.7	*<0.001
Yes	2 868	41.0	55.1		32.1	
Anaemia level						
Severe	224	3.7	87.6	*<0.001	67.4	*<0.001
Moderate	2 376	39.4	61.9		40.6	
Mild	1 524	25.3	37.5		20.6	
Not anaemic	1 905	31.6	25.1		12.9	
Total	7 011	100.0	41.5		27.3	

*RDT+: positive on the rapid diagnostic test, Micro+: positive on microscopy test.*significant at 5% x2 test*

The outcome of the McNemar paired test used in testing the hypothesis of non-agreement between the RDT and microscopy tests showed evidence of a significant agreement between the two tests with test statistics of 553.47 with p<0.0001 while the Kappa statistics of the agreement was 0.55 (not shown in
117 the Tables).

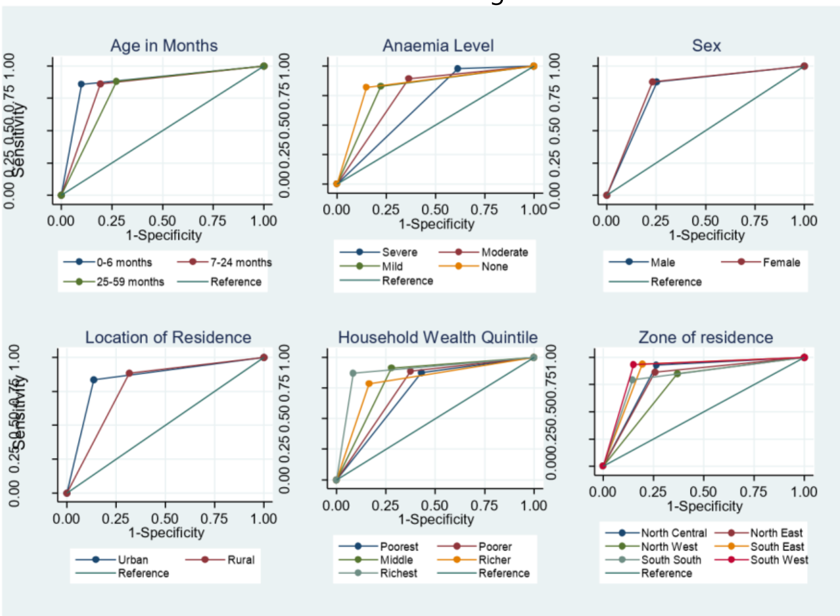
Based on the microscopy test as a gold standard, 118 Table 2 shows the discriminatory and predictive accuracies of the RDT test by the characteristics of the 5753 children that had both tests. The overall sensitivity was 88%, specificity was 76% while the diagnostic accuracy (that is the overall agreement) of RDT stood at 79%. The overall positive and negative likelihood ratio was 3.6 and 0.16 respectively while the overall PPV and NPV were 58% and 94% respectively. The sensitivity of RDT increased significantly as the age of the children increased ($p<0.05$), from 86% among those aged 0-6 months to 86.1% in 7-23 month olds to 88.1% among those aged 24-59 months. The reverse was the specificity, LR+, and the NPV but diagnostic accuracy of RDT and the PPV increased with the age of the children ($p<0.05$). On the sex of the children, there was no difference in the sensitivity of RDT but the specificity differed with significantly higher proportions among females (77%) than males (75%) ($p<0.05$). Interestingly, RDT was more sensitive in the rural area at 88% than in urban areas (84%) but less specific in rural (68%) than in the (86%) ($p<0.05$).

Also, there was a significant PPV gap of 14% in a rural area (60%) compared with the urban area (46%) ($p<0.05$) but with a closer NPV (92% vs 98%) ($p<0.05$). The overall accuracy, the sensitivity, the specificity, LR+, and NPV of the RDT were generally and significantly higher across the southern regions in Nigeria than in the northern regions ($p<0.05$). For children with severe anaemia, the sensitivity of the RDT was nearly 100% compared with a specificity of 39% ($p<0.05$). It appeared that the discriminatory accuracy reduced with children level of anaemia but the higher the severity of the anaemia, the lower its specificity ($p<0.05$). However, both the LR+ and the LR- increased as the severity of anaemia increased ($p<0.05$). For the predictive accuracy, PPV reduced with reduced severity from 77% to 44% ($p<0.05$) among children that didn't have anaemia while the NPV increased significantly from 90% among children that had severe anaemia to 97% among those that didn't have anaemia ($p<0.05$).

Among all the under-five children, the prior probability of malaria was 27%. For the positive tests (showed in blue line in Figure 1), the LR+ is 3.62 (95% CI: 3.42-3.83), the posterior probability (odds) is 58% (95% CI: 56%-59%) and approximately 1 in 1.7 with positive test results actually had the parasite while approximately 1 in 1.1 who tested negative was actually negative, LR- is 0.16(95% CI:0.14-0.19) with corresponding posterior probability (odds) of 0.06(95% CI: 0.05-0.07).

On the level of anaemia, AUC was 68% for those that had severe anaemia, 77% for moderate anaemia, 80% for mild anaemia and 84% for those with no anaemia ($p<0.05$). The AUCs were significantly different across the age of the children, the zone and location of their residence, their household wealth quintile, mothers' education, quality of housing material, haven had fever within two weeks preceding the data collection as well as the level of anaemia ($p<0.05$). The ROC and Lorenz curves for all children are shown in Supplementary Figure 1 while the ROC curves for some selected children characteristics are shown in Figure 1.

Figure 1: The ROC curve of the accuracy of the RDT test by the selected characteristic of under five children in Nigeria 2015 MIS



The adjusted odds of RDT being sensitive was about 50% (aOR=0.52; 95% CI: 0.30-0.90; $p<0.05$) lower among children aged 7-23 months compared with those aged 24-59 months. The adjusted odds of RDT being sensitive was 11 times (aOR=11.30; 95% CI: 2.58-49.60; $p<0.05$) and 2 times (aOR=2.15; 95% CI:1.67-2.77; $p<0.05$) higher among those aged 0-6 months and 7-23 months respectively than among those aged 24-59 months. Also, the adjusted odds of RDT being sensitive was significantly over five times more likely among children with severe anaemia than those with moderate anaemia, while the odds of being specific was about three times higher among children with no anaemia than children with moderate anaemia ($p<0.05$).

Table 2: Adjusted factors influencing the sensitivity, specificity and diagnostic accuracy of the RDT compared with the microscopy in Nigeria 2015 MIS

Characteristics	Sensitivity		Specificity		Diagnostic Accuracy	
	aOR (95% CI)	p-value	aOR (95% CI)	p-value	aOR (95% CI)	p-value
Child Age (24-59 Months)						
0-6 Months	0.53 (0.04-6.84)	0.63	11.3 (2.58-49.6)	0.00	3.88 (1.16-12.9)	0.03
7-23 Months	0.52 (0.30-0.90)	0.02	2.15 (1.67-2.77)	0.00	1.32 (1.07-1.62)	0.01
Anaemia Level (Moderate)						
Severe	5.46 (1.27-23.4)	0.02	0.61 (0.3-1.24)	0.17	1.84 (1.16-2.94)	0.01
Mild	0.80 (0.41-1.56)	0.52	1.80 (1.36-2.40)	0.00	1.13 (0.89-1.44)	0.31
Not Anaemic	0.38 (0.19-0.76)	0.01	2.88 (2.15-3.86)	0.00	1.45 (1.13-1.86)	0.00
Treat Malaria Wltw						
No	1.73 (1.26-2.38)	0.00	1.53 (0.76-3.08)	0.23	1.33 (0.9-1.98)	0.16
Zone (North West)						
North Central	6.49 (1.84-22.8)	0.00	0.8 (0.55-1.17)	0.25	0.95 (0.69-1.30)	0.74
North East	1.31 (0.69-2.52)	0.41	1.15 (0.83-1.59)	0.39	1.02 (0.78-1.32)	0.89
South East	2.52 (0.6-10.55)	0.21	0.87 (0.53-1.41)	0.57	1.0 0(0.66-1.52)	1.00
South South	1.45 (0.53-3.95)	0.47	1.3 (0.79-2.13)	0.30	1.21 (0.80-1.83)	0.36
South West	3.29 (0.82-13.25)	0.09	0.78 (0.44-1.35)	0.37	0.87 (0.54-1.40)	0.58
Location						
Urban	0.54 (0.23-1.27)	0.16	1.24 (0.90-1.70)	0.19	0.99 (0.74-1.31)	0.92
Female	1.40 (0.86-2.30)	0.18	1.01 (0.81-1.28)	0.91	1.06 (0.87-1.28)	0.58
Wealth Quintile (Poorest)						
Poorer	0.73 (0.38-1.40)	0.34	1.22 (0.85-1.74)	0.28	1.21 (0.91-1.59)	0.19
Middle	1.24 (0.42-3.68)	0.70	1.65 (1.08-2.53)	0.02	1.42 (1.01-2.03)	0.04
Richer	0.34 (0.10-1.13)	0.08	2.31 (1.37-3.91)	0.00	1.49 (0.95-2.32)	0.08
Richest	0.53 (0.10-2.73)	0.44	3.04 (1.61-5.71)	0.00	2.19 (1.26-3.83)	0.01

4. Discussion

This study was designed to evaluate the discriminatory and predictive accuracy of the RDT malaria test against the microscopy smear test as the "gold" standard. There were significant relationship and agreement between the outcomes of the RDT and microscopy malaria test procedures. However, the Kappa statistics at 0.55 and the Gini and Pietra indexes both at 0.6337 showed that the agreement was only good but not excellent. Generally, the RDT was found to be more sensitive (88%) than being specific (76%). The sensitivity in the current study is comparable to the sensitivity of 89% but lower than the specificity of 88% reported in an earlier longitudinal study in Tanzania [8]. In the same report, 63% sensitivity and 94% specificity were obtained in the cross-sectional component of the study. The authors noted that these measurements of discriminatory analysis varied widely and it depended on the presence of fever and the parasite density [8].

Murungi et al. noted that poor specificity in a diagnostic test may negatively impact RDT-based diagnostic strategies for malaria [27]. The wide disparities in the performance of RDTs against the microscopy and the PCR motivated Murungi et al to explore the accuracy of the HRP2 and pLDH RDTs and the microscopy in a two-step algorithm among 276 individuals. The authors found varying and very high levels of sensitivity and specificity depending on the stage of malaria. They concluded that certain RDTs could be more accurate in new cases and initial diagnosis than in malaria case monitoring and treatment and vice-versa [27].

On the predictive accuracy of the RDTs used in malaria diagnosis in the current study, the PPV was low at 58% while the NPV was relatively high at 94% respectively. This finding suggests that the likelihood of having malaria when the RDT test is positive is only about half while the likelihood of being disease-free when the RDT test is negative is very high. In contrast, Hopkins et al found that the PPV of the HRP2-based test was 98% compared with “the expert” microscopy, with an NPV of 97% for the HRP2-based test [15]. Also, a study carried out in Burkina Faso found the PPV and NPV of RDT to be 9% and 99.8% respectively in the dry season compared with 82% and 84% respectively in the rainy season for infants to over 99% for adults [14]. However, it is worth noting that PPVs are functions of disease prevalence.

Also, the level of anaemia influenced both the discriminatory and the predictive accuracies of the RDTs. The RDTs were totally sensitive and less than 40% specific among children with severe anaemia. The higher the severity of anaemia in children, the higher the sensitivity and the lower the specificity. In a similar trend, PPV reduced with reducing anaemia severity while the NPV increased with a reduction in the level of severity of anaemia.

Although, all the tests in the current study were carried out during the same dry season which eliminated seasonal variability. It cannot be ascertained in the current study if the dry season influenced the accuracies of the RDTs in a study conducted in Burkina Faso. Bisofi et al. had found a significant effect of seasonality in the discriminatory accuracy of RDT. It was reported that while the sensitivity and specificity of the RDT were 86% and 90% respectively in the dry season, the figures were 94% and 78% respectively in the rainy season [14]. The same study found seasonal variability in PPV and NPV of RDT to be 9% and 99.8% respectively in the dry season compared with 82% and 84% respectively in the rainy season among infants [14]. In addition, Mouatcho et al. found that the specificities, sensitivities, numbers of false positives, numbers of false negatives and temperature tolerances of the RDTs vary considerably and are some of the challenges facing the accuracy of RDTs [16]. Other factors that may influence diagnostic accuracies are the efficiency of RDT storage, transport or handling of malaria RDTs as well as the expertise of the handlers [7] but these factors are not available for assessment in the current study.

Malaria RDTs are generally designed to be used in malaria-endemic areas where good-quality microscopies are out of reach. It is a requirement that a diagnostic method must be accurate and available at the point of care if the effective diagnosis of all malaria cases is to be achieved. A diagnostic test with a high degree of sensitivity usually has a low false negative rate which ensures that only a few true cases are not correctly classified. It is therefore imperative that a screening test used in ruling out cases should have a reasonably high degree of sensitivity. In the same vein, a diagnostic test with a high degree of specificity produces low false positive rates which invariably leads to only a few misdiagnosed subjects. A confirmatory test used in ruling-in cases should have a high degree of specificity.

There is a need to adhere to the WHO recommendations on the procurement of malaria RDTs [7] as well as on the storage and use of malaria RDTs. Total adherence to these recommendations will substantially improve the discriminatory and predictive accuracy of the malaria RDTS. According to the WHO, to be reliable, "RDTs should be sensitive enough to reliably detect malaria parasites at densities associated with disease" [28]. This suggests that sensitivity of an RDT is a function of the quality of manufacture, species, number, viability, and strain of parasites present, RDT conditions, storage conditions, application technique and level of care exercised. The WHO recommendation stated further that choice of RDT must be guided by the panel detection score (PDS) against the *Plasmodium falciparum* (Pf) and against the *Plasmodium vivax* (Pv) which must be at least 75% in both cases and that the false positive rates and invalid rate should be less than 10% and 5% respectively [7]. For RDTs to perform optimally, there must be a demonstration of the presence of parasitemia, efficient mechanism for quality control, "cool chain" for transport and storage, adequately trained health worker and, adequate monitoring [7].

Does Google search index help track and predict inflation rate? An exploratory analysis for India

G. P. Samanta[1]

Abstract

The forward looking outlook or market expectations on inflation constitute valuable input to monetary policy, particularly in the 'inflation targeting' regime. However, prediction or quantification of market expectations is a challenging task. The time lag in the publication of official statistics further aggravates the complexity of the issue. One way of dealing with non-availability of relevant data in realtime basis involves assessing the current or nowcasting the inflation based on a suitable model using past or present data on related variables. The forecast may be generated by extrapolating the model. Any error in the assessment of the current inflationary pressure thus may lead to erroneous forecasts if the latter is conditional upon the former. Market expectations may also be quantified by conducting suitable surveys. However, surveys are associated with substantial cost and resource implications, in addition to facing certain conceptual and operational challenges in terms of representativeness of the sample, estimation techniques, and so on. As a potential alternative to address this issue, recent literature is examining if the information content of the vast Google trend data generated through the volume of searches people make on the keyword 'inflation' or a suitable combination of keywords. The empirical literature on the issue is mostly exploratory in nature and has reported a few promising results. Inspired by this line of works, we have examined if the search volume on the keywords 'inflation' or 'price'in the Google search engine is useful to track and predict inflation rate in India. Empirical results are very encouraging. Future research may focus on fine-tuning of the present work further and to check the robustness of the results over time and across countries.

Keywords

Inflation Expectations; Surveys; Internet Search; Google Trend; Google search Index

[1] G.P.Samanta is an Adviser in the Department of Statistics and Information Management (DSIM), Reserve Bank of India (RBI). The results presented in the paper and the views expressed are purely personal and may not represent those of the institute/organization he belongs to. Comments and suggestions are welcomed. The authour can be reached through email at gpsamanta@rbi.org.in.

1. Introduction

Inflation expectations constitute an important ingredient to monetary policy formulation, particularly under the Inflation Targeting approach. However, the forward-looking assessment or forecasting inflation has been an extremely challenging task. The time lag in the release of official statistics on inflation often aggravates the problem further. To address the issue, the conventional literature suggests two broad approaches, viz., developing forecasting models and conducting surveys for measuring inflation expectations. The modelling exercise usually attempts to exploit inter-relationship between inflation and relevant economic variable and indicators, either under some economic theories or extracting data-driven patterns. The empirical estimation of such models uses data released by official statistics, which are traditionally compiled offline at a fixed interval, and usually released by compiling agencies with a substantial lag. At times, it may also employ the survey-based results as additional information.

Surveys have been an alternative tool to quantify market expectation on an economic variable/parameter and also to fill-in potential data gap or providing nowcasting indicator for a target variable, mainly data on which are released with lag. As regards surveys for inflation sentiments, international best practices have devoted to assessing future inflation based on either qualitative or quantitative responses from welldesigned and representative target group. While qualitative responses help in assessing the direction of change or movement of the future inflation rate, quantitative results directly reflect the respondents' perception about the level of future inflation numerically. Answering qualitative questions on a variable, mostly asking an opinion on ‘no change' or direction of change of the variable, are easier than providing a quantitative response, and may improve the response rate in a survey. There has been a quite rich literature on converting the qualitative responses on a variable to the corresponding numeric value of the variable. However, conducting surveys may be costly in terms of monetary expenditure, time requirement and human resources. Further, for the time requirement, survey-based results may fail to capture information real-time basis. To address these issues, many researchers have explored if vast metadata and documents available freely in online resources could be useful in tracking and predicting economic variables. With the advent of the internet and digital platform fast growing habits of people on the internet searching, expressing opinions and sentiments in social media and digital platform, a few researchers have examined if the internet resources available at more timely and more frequent manner than traditional data can be useful to assess expectations of economic agents. There have been wide varieties of online resources such as websites of business houses, online retailers, social media like Twitter, search queries in internet search engine like Google, digital or printed documents uploaded by various statistical and Government agencies, academic institutes, policy makers and regulators, etc. Each of these alternative resources has been experimentally assessed by the researchers for different purposes. For example, Choi and Varian (2009a, 2009b, 2012), Ettredge, et al. (2005), Guzmán, 2011, and Seabold and Coppola (2015) explored the usefulness of Google search data on tracking/nowcasting various economic activities and macroeconomic variables; Agarwal et al. (2011) attempted sentiment analyses based on Twitter messages; Cavallo (2013, 2015, 2016 & 2017), and Cavallo and Rigobon (2011, 2016) constructed price indices based on online prices and analysed various aspects of prices; and Cavallo et al. (2015) have studied the price impact of joining a currency union.

India has a long tradition of model building for analysing and predicting the inflation rate. In addition to building macro models and simultaneous-equation based system, researchers usually exploited multiple alternative techniques with varying degree of success. First, univariate time series models – both linear and non-linear – and both conditional homoscedastic and heteroscedastic models. This approach models interrelationship between current and past observations of time series data on a target variable. Second, single equation models by regressing inflation on own past as well as present and past observations of influential variables or determinants of inflation. Third, multivariate time series models exploiting the interrelationships of inflation and one or more related variables. These models could be either pure datadriven, such as Vector Auto Regression (VAR) or model developed following certain economic principles as could be done in structural-VAR or Vector Error Correction Models (VECM) under the co-integration framework. Fourth, the construction of composite leading indicators for tracking Inflation. Fifth, estimating economic-theory based models, such as P-Star model, different variants of Phillips-Curve or output-gap models. For past several years, the Reserve Bank of India (RBI), India's Central Bank, has been conducting many monetary policy surveys to gauge market expectations on inflation, growth and other economic parameters. While some of these surveys capture qualitative responses from the respondents, a few are capturing quantitative forecasts. As regards inflation, two of these surveys, viz., 'Inflation Expectation Survey of Household' (IESH), and 'Survey of Professional Forecasters' (SPF) capture quantitative forecasts of inflation. While the target group of respondents to IESH covers households, the SPF, as the name suggests, captures responses from a select list of professional forecasters[2].

Though inflation in India has modelled by various approaches in the past, hardly any attempt is made to assess the information content of online

[2] The SPF captures forecasts for inflation along with many other macroeconomic variables, such as growth, export, import, the exchange rate from professional forecasters.

resources to track inflation behaviour. Accordingly, this paper examines empirically if the data on internet search queries using Google engine can be gainfully employed to predict inflation for India. The plan of the paper is as follows. Section 2 provides a literature review on predicting inflation based on Google search index and the consumer-theory based approach adopted by Guzmán (2011). Section 3 presents data and empirical results, and Section 4 concludes.

2. Methodology

As this paper focuses on assessing the information content of Google search index for inflation rate, the dataset consists of two main components: First, the official statistics on price index which forms the basis of estimating inflation. Second, the Google search index for suitable keywords. Monthly data on these series are collected for a period of seven years from April 2012 to March 2019.

2.1 Data on Price Index

In India, inflation is now measured by annual percentage changes in Consumer Price Index (CPI) compiled by National Statistics Office (NSO), Ministry of Statistics and Programme Implementation (MoSPI), Government of India. Monthly data are released under three broad heads, viz., CPI-Urban (CPI-U), CPIRural (CPI-R) and CPI-Combined (CPI-C). While CPI-R and CPI-U represent price index for rural and urban India, respectively, the CPI-C is overall price index arrived at by combining CPI-R and CPI-U.

2.2 Google Search Indicators/Indices

Each Google trend series is characterised by two important features (Guzmán, 2011; Seabold and Coppola 2015): First, the numbers at various time points over the data period do not provide the absolute search volumes on the given keywords. Instead, they represent relative estimates in a sense that the time point with maximum search interest over the entire enquiry period is assigned a value 100 and the actual search volumes in other time points are rescaled accordingly. Second, the time series replica of search index on given key words for a specified period depends on the date when the search enquiry was made. Thus, time series data may change with search date even when the reference period for search index remain unchanged. These typical issues with Google search index have been handled by a two-step process. We first gather replicas of time series data on Google search Index on the keyword 'inflation' (with location: India) for the period April 2012 to March 2019 on three different dates of May 2019. As expected, the replica of relative measures on three different dates appears numerically different. We constructed an overall replica of time series by taking geometric-mean of the replicas obtained in three different dates and denote these geometric-mean based overall indices for the keywords 'inflation' and 'price' as GMInfl and GMPrice, respectively.

2.3 Change in Indices/Variables, Inflation Rates and Codes/Symbols Used for Different Variables

The k-period inflation rates for t-th month, based on CPIs are the log-return or continuously compounding, are computed as

$$\pi_t^m = 100 * [\ \log_e X_t - \log_e X_{t-m}\ ] \qquad ....3$$

Where π_t^m is the m-period change or inflation rate at t-th month; m=1 for monthly change/inflation and m=12 for annual inflation/change; X=CPI-C, CPI-U, CPI-R or Google trend indices GMPrice, GMInfl.

For convenience in referencing any transformation or derived variable for a time series, say X, we used codes/symbols as follows; gX = annual percentage changes as in Eqn. (3); lnX = $\log_e(X)$; ΔX and Δ2X represent 1st and 2nd difference of X, respectively; eX and e2X are residual obtained by fitting linear and quadratic time trends on X respectively; where X = CPI-C, CPI-U, GMPrice, GMInfl or corresponding logtransformations.

3. Results

The assessment of the information content of a given Google search Index is carried out in a multi-step process. First, we examine basic time series properties, such as stationarity or non-stationarity, of offline price indices, Google trend index, change in indices and inflation rates.

3.1 Tests for Unit-Root – Difference-Stationary and Trend-Stationary Processes

We applied Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests for examining if (a) log-transformed indices are stationary, and (b) if any detected non-stationary series belongs to trend-stationary (TS) or difference-stationary (DS) class. The equation considered for implementing ADF tests for a time series Xt is of the following general form.

$$\Delta X_t = \alpha + \beta t + \rho X_{t-1} + \sum_{i=1}^{1} \delta_i \Delta X_{t-i} + \varepsilon_t \qquad ......(4)$$

Where Δ is the difference operator; α, β, ρ and δ_i are unknown constants, and ε_t is the usual error series.

In Eqn. (4), parameters of interest are β and ρ. If β=0 and ρ<1 then X_t is a stationary, i.e. I(0) series, and if (β,ρ) =(0,1) then X_t has unit-root and belongs to difference-stationary process, i.e. X_t is non-stationary I(1) and ΔX_t is I(0) process. However, if β ≠ 0 and ρ < 1, then also X_t is non-stationary and belongs to trend-stationary (TS) series. Removal of deterministic time trend from a TS series would yield a stationary or I(0) series. The unit-root tests were first carried out for for log(CPI-C), log(CPI-U), log(GMPrice) and log(GMInfl) directly, and found some mixed results (Table 1). It appears that while lnCPI-C and lnCPI-U are I(2) processes, lnGPPrice and lnGMInfl are I(1) processes. Further, we examined the stationarity or unit-root properties of annual inflation rates based on CPI-C and CPI-U, and the annual percentage change in GMPrice and GMInfl computed in line with Eqn. (3). The test results (Panel A of Table 2), identified all these annual inflation/change series to be I(1). As a reconfirmation to these findings, the unit-root tests identified the first-difference of each of the four annual change series to be stationary, i.e. I(0) processes (Panel B, Table 2).

Table 2: Unit-Root Tests – Different transformed-variables

Variable	Augmented Dickey-Fuller					Phillips - Perron				
	Optimal Lag	Unit-Root Test		Test for Trend		Band-width	Unit-Root Test		Test for Trend	
		Test Statistics	p-value	Test Statistics	p-value		Test Statistics	p-value	Test Statistics	p-value
(A) Annual Inflation Rate/Annual Percentage Change										
lnCPI-C	7	-2.8429	0.1869	1.7515	0.0843	3	-2.2117	0.4769	1.3288	0.1876
lnCPI-U	7	-3.7490	0.0247	2.8582	0.0056	3	-2.6516	0.2592	1.9785	0.0512
lnGMPrice	6	- 3.9981	0.0125	3.0147	0.0036	1	-3.4091	0.0569	1.8850	0.0629
lnGMInfl	0	-5.8940	0.0000	-5.2213	0.0000	2	-5.8860	0.0000	-5.2213	0.0000
(B) First-Difference of Variable at (A) above										
ΔlnCPI-C	6	-7.0514	0.0000	-4.3018	0.0001	0	-5.6672	0.0000	-2.0093	0.0478
ΔlnCPI-U	6	-6.4774	0.0000	-3.4508	0.0010	2	-6.0054	0.0000	-1.7980	0.0759
ΔlnGMPrice	3	-5.9215	0.0000	-0.9223	0.3593	8	-8.1457	0.0000	-0.6333	0.5283
ΔlnGMInfl	1	-9.4089	0.0000	-0.0126	0.9900	13	-20.5903	0.0000	0.1838	0.8546
(C) Second-Difference of Variables at (A) above										
Δ^2lnCPI-C	7	-7.8157	0.0000	0.3484	0.7287	8	-15.7029	0.0000	0.1071	0.9150
Δ^2lnCPI-U	7	-7.7350	0.0000	0.5843	0.5610	4	-11.9371	0.0000	0.2176	0.8283
Δ^2lnGMPrice	4	-8.7383	0.0000	-0.0139	0.9889	6	-51.1038	0.0001	0.0135	0.9892
Δ^2lnGMInfl	4	-7.7786	0.0000	-0.0493	0.9913	6	-51.2341	0.0001	0.0234	0.9852
(D) De-Trending Linear -Time-Trend of variables at (A) above										
elnCPI-C	7	-2.8428	0.1869	-5.1194	0.0000	1	-5.3040	0.0002	0.0916	0.9273
elnCPI-U	7	-3.7490	0.0247	-4.8328	0.0000	3	-2.6516	0.2592	-2.9244	0.0044
elnGMPrice	6	-3.9981	0.0125	-0.5160	0.6075	1	-3.4091	0.0569	-0.6357	0.5267
elnGMInfl	0	-5.8940	0.0000	0.1911	0.8489	2	-5.8860	0.0000	0.1911	0.8489
(E) De-Trending Quadratic-Time-Trend of Variables at (A) above										
e2lnCPI-C	1	-4.4954	0.0027	-0.5439	0.5880	3	-3.2540	0.0811	-0.4989	0.6191
e2lnCPI-U	1	-3.6238	0.0336	-0.4053	0.6864	3	-2.8794	0.1743	-0.3621	0.7182
e2lnGMPrice	3	-3.6125	0.0348	-0.4351	0.6647	1	-3.5711	0.0384	-0.3378	0.7364
e2lnGMInfl	0	-6.0223	0.0000	0.0601	0.9522	1	-6.0519	0.0601	0.0601	0.9522

Table 2: Unit-Root Tests – Annual Inflation Rate/Change in Google Index

Variable	Augmented Dickey-Fuller					Phillips - Perron				
	Optimal Lag	Unit-Root Test		Test for Trend		Band-width	Unit-Root Test		Test for Trend	
		Test Statistics	p-value	Test Statistics	p-value		Test Statistics	p-value	Test Statistics	p-value
(A) Annual Inflation Rate/Annual Percentage Change										
gCPI-C	1	-3.1547	0.1018	-2.1941	0.0316	3	-2.3619	0.3961	-1.4354	0.1556
gCPI-U	4	-1.8559	0.6669	-0.7415	0.4610	5	-1.8075	0.6912	-0.5764	0.5662
gGMPrice	0	-5.3040	0.0002	0.0916	0.9273	1	-5.3040	0.0002	0.0916	0.9273
gGMInfl	0	-2.2835	0.4373	-0.8676	0.3885	4	-2.5811	0.2900	-0.8676	0.3885
(B) First-Difference Series of Variable at (A) above										
ΔgCPI-C	11	-4.3419	0.0052	1.5623	0.1248	13	-6.2063	0.0000	0.7070	0.4819
ΔgCPI-U	1	-6.6947	0.0000	1.0733	0.2869	13	-5.9064	0.0000	1.0946	0.2774
ΔgGMPrice	1	-9.1522	0.0000	0.1058	0.9161	26	-20.2929	0.0001	-0.1534	0.8785
ΔgGMInfl	0	-7.8708	0.0000	-0.2613	0.7946	2	-7.8397	0.0000	-0.2613	0.7946

3.2 Predictive Ability - Granger Causality Tests

The predictive or forecasting ability of Google trend data is assessed under the Granger Causality framework (Guzmán (2011). This technique is also useful to test if past price situation or realised inflations have any bearing on volume of internet search. The Granger causality tests for a pair of variables, say X_t and Y_t are carried out based on following general equations.

$$X_t = \alpha_0 + \sum_{i=1}^{l} \alpha_i X_{t-i} + \sum_{j=1}^{l} \beta_j Y_{t-j} + \varepsilon_t \qquad \text{..... (5)}$$

$$Y_t = \alpha_0 + \sum_{i=1}^{l} \alpha_i X_{t-i} + \sum_{j=1}^{l} \beta_j Y_{t-j} + \varepsilon_t \qquad \text{.......(6)}$$

Where, α_i's, i=0,1, ... and β_j's, j=1,2,.... are unknown constants; 1 is suitable chosen positive integer; and ε_t is usual residual/error series.

We examine causal relationship between some form of inflation or transformed price indices and Google search indices for relevant search words. In particular, we considered two price indices, viz., CPI-C and CPI-U, and two Google search indicators, viz., GMPrice and GMInfl.

Table 5: Predictive Power – Granger Causality

Google Search Data	**Null Hypothesis**	**Obs**	**Lag**	**F-Statistics**	**P-Value**
gGMPrice	gGMPrice does not Granger Cause gCPI-C	66	9	1.9125	0.0732
	gCPI-C does not Granger cause gGMPrice	66	9	2.0230	0.0575
	gGMPrice does not Granger Cause gCPI-U	72	3	0.6434	0.5899
	gCPI-U does not Granger cause gGMPrice	72	3	3.3392	0.0246

ΔgGMPrice	ΔgGMPrice does not Granger Cause ΔgCPI-C	72	2	0.9296	0.3937
	ΔgCPI-C does not Granger cause ΔgGMPrice	72	2	3.4142	0.0387
	ΔgGMPrice does not Granger Cause ΔgCPI-U	72	2	1.7898	0.1749
	ΔgCPI-U does not Granger cause ΔgGMPrice	72	2	3.3384	0.0415
ΔgGMInfl	ΔgGMInfl does not Granger Cause ΔgCPI-C	62	12	1.6213	0.1279
	ΔgCPI-C does not Granger cause ΔgGMInfl	62	12	2.0394	0.0483
	ΔgGMInfl does not Granger Cause ΔgCPI-U	62	12	1.6005	0.1341
	ΔgCPI-U does not Granger cause ΔgGMInfl	62	12	2.8420	0.0073
Δ2lnGMInfl	Δ2lnGMInfl does not Granger Cause Δ2lnCPI-C	66	7	2.0053	0.0721
	Δ2lnCPI-C does not Granger cause Δ2lnGMInfl	66	7	2.2713	0.0431
	Δ2lnGMInfl does not Granger Cause Δ2lnCPI-U	63	10	1.4382	0.1974
	Δ2lnCPI-U does not Granger cause Δ2lnGMInfl	63	10	2.0800	0.0483
e2lnGMPrice	e2lnGMPrice does not Granger Cause e2lnCPI-C	78	9	2.7336	0.0098
	e2lnCPI-C does not Granger cause e2lnGMPrice	78	9	2.6099	0.0131
	e2lnGMPrice does not Granger Cause e2lnCPI-U	78	9	2.8349	0.0077
	e2lnCPI-U does not Granger cause e2lnGMPrice	78	9	3.4623	0.0017
e2lnGMInfl	e2lnGMInfl does not Granger Cause e2lnCPI-C	78	9	1.1596	0.3334
	e2lnCPI-C does not Granger cause e2lnGMInfl	78	9	2.0825	0.0456
	e2lnGMInfl does not Granger Cause e2lnCPI-U	78	9	0.6332	0.7640
	e2lnCPI-U does not Granger cause e2lnGMInfl	78	9	2.2178	0.0332

3.3 Correlation

The tracking ability of GMPrice or GMInfl is examined simply by correlation coefficient between GMPrice/GMInfl and price index either in their original or stationary-transformed forms. Significance of these correlation coefficients would establish the tracking or nowcasting ability of Google search index. Table 4 presents correlation coefficients for (a) different pairs of annual rate of inflation or change in CPIC, CPI-U, GMPrice and GMInfl; (b) monthly change (i.e. first difference) of the annual inflation or growth rates; and (c) similar results for stationary-transformed series for lnCPI-C, lnCPI-U, lnGMPrice, and lnGMInfl. As seen, the annual percentage change in GMInfl is strongly correlated with annual inflation rates based on both CPI-C and CPI-U, indicating that Google search index GMInfl is useful in tracking annual inflation rates. Further, positive sign of correlation coefficients indicates the pairs of variables usually movement in same direction.

Table 4: Correlation Coefficient between Different Pairs of Variables

Variable Pair	**Correlation Coefficient**		**Variable Pair**	**Correlation Coefficient**	
gGMPrice & gCPI-C	0.1805	(0.1212)	gGMInfl & gCPI-C	0.3591	(0.0016)
gGMPrice & gCPI-U	0.1641	(0.1595)	gGMInfl & gCPI-U	0.3509	(0.0020)
ΔgGMPrice & ΔgCPI-C	0.0150	(0.8985)	ΔgGMInfl & ΔgCPI-C	0.1608	(0.1704)
ΔgGMPrice & ΔgCPI-U	-0.0082	(0.9447)	ΔgGMInfl & ΔgCPI-U	0.1710	(0.1450)
ΔlnGMPrice & ΔlnCPI-C	0.1695	(0.1186)	ΔlnGMInfl & ΔlnCPI-C	0.0373	(0.7333)
ΔlnGMPrice & ΔlnCPI-U	0.1401	(0.1981)	ΔlnGMInfl & ΔlnCPI-U	-0.0280	(0.7977)
e2lnGMPrice & e2lnCPI-C	0.3977	(0.0001)	e2lnGMInfl & e2lnCPI-C	0.2557	(0.0168)
e2lnGMPrice & e2lnCPI-U	0.3701	(0.0013)	e2lnGMInfl & e2lnCPI-U	0.2913	(0.0062)

Figures within ( ) are p-values.

4. Discussion and Conclusion

Prediction of inflation or quantification of market expectations on inflation is always very challenging. Other than model-based forecast, a conventional way of measuring market expectations involves conducting suitable surveys to capture current assessment and forward-looking outlook of the appropriate target group. However, surveys are associated with substantial cost and resource implications, in addition to facing specific conceptual and operational challenges. As a potential alternative to address the issue, many researchers argue that interest on Google search on a relevant keyword, such as 'inflation' reflects the 'revealed expectations' of people and examine if the Google search volume track or predict inflation rate. Though the short empirical literature on the subject is mainly exploratory in nature, some of those studies have reported quite encouraging results.

In this paper, we assess the information content of Google search volume on two relevant keywords, viz., 'price' and 'inflation' in tracking or predicting the inflation rate in India. Empirical results show that such an index for the keyword 'inflation' is useful to track inflation rates India based on both CPI-Combined and CPI-Urban. Granger's causality tests also detect the strong predictive ability of the search index. Future research in this emerging area can be generalised in various ways, such as examining the information content of Google search data about related keywords, checking the robustness of the findings at different sub-national regions of India.

References

1. Agarwal, Aprrov, Boyi Xie, Ilia Vovsha, Owen Rambow and Rebecca Passonneau (2011), "Sentiment Analysis of Twitter Data", ***Proceedings of the Workshop on Language in Social Media (LSM 2011), pages 30–38, Portland, Oregon***, 23 June 2011.
2. Cavallo, Alberto (2013), "Online and Official Price Indexes: Measuring Argentina's Inflation", ***Journal of Monetary Economics***, Vol. 60, pp. 152-65.
3. Cavallo, Alberto (2015), "Scraped Data and Sticky Prices", ***NBER Working Paper Series***, Working Paper 21490.
4. Cavallo, Alberto (2016), "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers", ***NBER Working Paper Series***, Working Paper 22142, March.
5. Cavallo, Alberto (2017), "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers", ***American Economic Review***, Vol. 107, No. 1, pp. 283-303.
6. Cavallo, Alberto, Brent Neiman and Roberto Rigobon (2015), "The Price Impact of Joining a Currency Union: Evidence from Latvia", ***IMF Economic Review***, Vol 63, No. 2.
7. Cavallo, Alberto and Roberto Rigobon (2011), "The Distribution of the Size of Price Changes", ***NBER Working Paper Series,*** Working Paper 16760.
8. Cavallo, Alberto and Roberto Rigobon (2016), "The Billion Prices Project: Using Online Prices for Measurement and Research", ***Journal of Economic Perspectives***, Vol. 30, No. 2, Spring, pp. 151-78.
9. Choi, Hyunyoung and Hal Varian (2009a), "Predicting the Present with Google Trends", ***Technical Report***, Google Inc. (**Website:** https://research.google.com/googleblogs/pdfs/ google_predicting_the_present.pdf).
10. Choi, Hyunyoung and Hal Varian (2009b), "Predicting Initial Claims for Unemployment Insurance Using Google Trends", ***Technical Report***,

Google Inc. (**Website**: https://research.google.com/archive/papers/initialclaimsUS.pdf).

11. Choi, Hyunyoung and Hal Varian (2012), "Predicting the Present with Google Trends", ***Economic Record***, ***The Economic Society of Australia***, Vol. 88, Issue S1 (Special Issue), pp 2-9.
12. Ettredge, Michael, John Gerdes and Gilbert Karuga (2005), "Using Web-based Search Data to Predict Macroeconomic Statistics", ***Communications of the ACM***, Vol 48, No. 11, November, pp. 87-92.
13. Ellen Tobback, Stefano Nardelli, David Martens (2017), "Between Hawks and Doves: Measuring Central Bank Communication", **Working Paper Series**, **European Central Bank**, No. 2086, July.
14. Guzmán, Giselle (2011), "Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations", ***Journal of Economic and Social Measurement***, Vol. 36, No. 3, IOS Press, pp. 119-167.
15. Kharde, Vishal A. and S.S.Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques", ***International Journal of Computer Applications***, Vol. 139, No. 11, April, pp. 5-15.
16. Mika V. Mäntylä, Daniel Graziotin, Miikka Kuutila (2018), "The Evolution of Sentiment Analysis—A Review of Research Topics, Venues, and Top Cited Papers", **Computer Science Review**, Volume 27, February 2018, pp. 16-32.
17. Seabold, Skipper and Andrea Coppola (2015), "Nowcasting Prices Using Google Trends – An Application to Central America", ***Policy Research Working Paper 7398, World Bank Group***, August.

The role of big data and surveys in measuring and predicting inflation

Bruno Tissot
Head of Statistics and Research Support, Bank for International Settlements (BIS), Basel, Switzerland, and Head of the Secretariat of the Irving Fisher Committee on Central Bank Statistics (IFC)[1]

Abstract
Monitoring and forecasting short-term price developments is an important issue for public authorities. Measuring inflation requires significant resources, and substantial methodological work has been developed over the past decades to support this task. The recent emergence of Big Data can provide many opportunities in this context, for instance to produce timelier indicators, enhance the collection of specific types of prices, take due consideration of economic agents' expectations, and facilitate shortterm forecasting. However, there are important challenges when using big data-type of information, and traditional statistical surveys have continued to prove their usefulness. This suggests that measuring and forecasting inflation should continue to benefit from drawing on multiple, complementary approaches.

Keywords
Prices; Public policy; Internet; Inflation expectations; Nowcasting

1. Introduction

Measuring and predicting the evolution of prices is key in a market economy, and it has therefore been traditionally an important objective for authorities. Indeed, past centuries were characterised by various episodes of strong money creation, leading to peaks in inflation with often important social and political consequences (see Graph 1 for an historical perspective on European inflation). Episodes of price deflation have been also particularly disruptive, and the Great Depression of the 1930s underlined the need for **properly measuring and anticipating both the real economic activity and the evolution of prices.** This was clearly a key factor driving the subsequent development of the Systems of the National Accounts (SNA; European Commission et al (2009)) framework. Today, most advanced economies and a growing number of developing economies have reliable statistics to measure consumer price inflation (CPI) based on detailed international guidance (ILO et al (2004)). Such indicators are often based on statistical,

[1] Assistance from Anamaria Illes and Antonio Perrella with the preparation of the graphs is gratefully acknowledged.

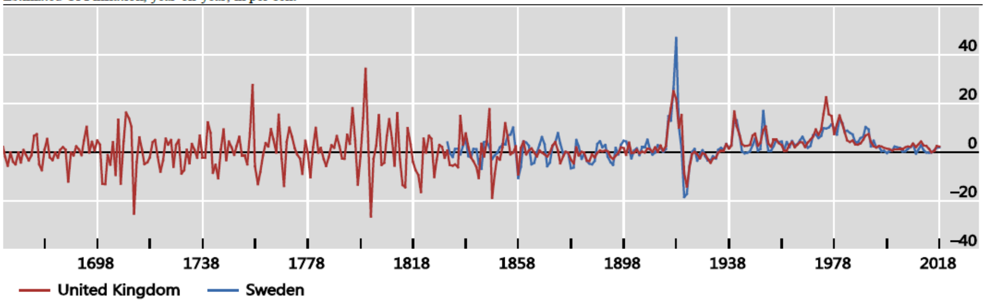
census-type surveys to collect the prices of a list of products in selected retail points and aggregated with proper weights to reflect the composition of households' consumption basket.

Nevertheless, **important limitations** still hinder a timely and comprehensive provision of reliable inflation indicators across the globe. Five key issues are worth highlighting from this perspective.

The history of inflation – selected European countries

Graph 1

Estimated CPI inflation, year-on-year; in per cent



Source: BIS.

First, a **robust statistical infrastructure** is required to produce inflation data on a regular basis; this calls for sufficient staff resources, adequate statistical skills, effective IT support, the set-up of specific processes to capture prices observed in various market segments, etc.

Second, the **impact of innovation and digitalisation** is posing practical and theoretical difficulties as regards the definition and the measurement of inflation (see Reinsdorf and P Schreyer (2019)). In particular, it is increasingly difficult to capture instant prices offered during limited periods of time and/or for specific groups of buyers. Moreover, correctly measuring inflation requires the ability to capture, and correct for, quality changes. Yet, increased innovation has reduced the ability to observe the repeated sales of the same products, since their characteristics are constantly evolving, hence making difficult the observation of prices for constant quality goods. One solution is to apply the so-called hedonic method.[2] But this requires capturing a wealth of product characteristics that have to be analysed using econometric techniques – making the data collection and compilation process more complex.

Third, inflation is a **multiform phenomenon** that varies across sectors and economic agents, leading to a multiplicity of inflation indicators. In general, one will refer to inflation as the evolution in the prices of goods and services consumed by an average consumer, ie the CPI index. However,

[2] Defined as a "regression technique used to estimate the prices of qualities or models that are not available on the market in particular periods, but whose prices in those periods are needed in order to be able to construct price relatives"; cf OECD Glossary of Statistical Term, available at stats.oecd.org/glossary/index.htm.

inflation will be perceived differently by other types of agents, for instance by producers, exporters or importers, or even within these groups themselves – say, between wealthy and poor consumers; but this type of distribution information is generally missing in the SNA framework (Tissot (2018)). Moreover, one could be more interested in the evolution of some sub-components of inflation – for instance in "core" inflation, which corrects for the volatility caused by movements in food and energy prices to reveal underlying inflation trends. The situation in the United States illustrates this variety of measures in the case of consumer inflation: the Bureau of Labour Statistics (BLS) produces monthly data on changes in the prices paid by urban consumers for a representative basket of goods and services, with separate indexes for two groups: all urban consumers (CPI-U), which is the index most often reported; and the urban wage earners and clerical workers (CPI-W), which is used more for wage negotiations (BLS (2019)). Several additional indicators are compiled, such as: the BLS chained CPI (designed as a "costof-living" index); the Personal Consumption Expenditures (PCE) price index developed by the Bureau of Economic Analysis in consistency with the SNA framework; the core PCE; etc.

Fourth, an important factor driving price and wage developments relates to **expectations**. To address this point, many countries have set up a number of surveys to help predict short-term developments in inflation – eg in the context of the work organised since 1953 under the umbrella of the Centre for International Research on Economic Tendency Surveys (CIRET), a global forum for institutions conducting business and/or consumer surveys. Such surveys comprise the well-known US consumer survey by the University of Michigan (Curtin (1996)), the business and consumer surveys conducted under the aegis of the European Commission in Europe (European Commission (2019)), and the consensus forecasts collected from various economists and forecasters – cf for instance the short- and long-term inflation forecasts of the US Survey of Professional Forecasters published by the Federal Reserve Bank of Philadelphia (Croushore (1993)) as well as the similar survey set up by the ECB (Garcia (2003)). Yet how useful such measures of inflation expectations can be to help predict actual inflation numbers is an open issue, as suggested by Graph 2 for the Euro Area and the United States.

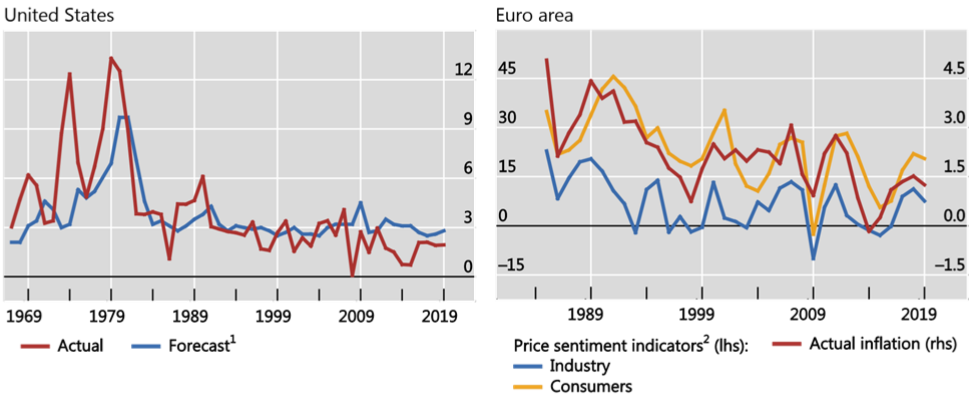
Fifth, inflation is a general concept that is not limited to traded goods and services. One important area relates to **asset prices**, which can play a powerful role as was seen during the Great Financial Crisis of 2007-09, with the strong impact of the ups and downs observed in housing prices (Tissot (2014)). In addition, developments in financial prices can also be useful to measure inflation expectations based on market-based indicators (eg index-linked bonds).

2. Opportunities provided by Big Data

Given the challenges mentioned above, what are the opportunities offered by big data – described by some as the new oil of the 21st century (The Economist (2017))? The main sources of big data are social networks, traditional business systems, and the internet of things,[3] and **four types of data sets** appear of particular importance in the economic and financial area: internet-based indicators, commercial data sets, financial market indicators and administrative records (IFC (2017)).[4] These data are often available in an "organic" way, unlike statistical surveys and censuses: the reason is that they are usually not collected ("designed") for a specific statistical purpose, being the by-product of other activities (Groves (2011)). Hence, there is a clear interest to use them, considering that the cost of launching traditional statistical surveys can be significant. Moreover, the new type of information they provide can help addressing some of the main challenges faced when measuring and predicting inflation.

Inflation expectations and actual observations
Year-on-year; in per cent

Graph 2



[1] Inflation rate changes expected one year earlier. [2] Annual averages, balances (%) of the price expectations for the next 12 months questions; not seasonally adjusted.
Sources: European Commission; University of Michigan; national data; BIS calculations.

As regards the first issue related to the need for a robust statistical infrastructure, big data can **facilitate the work of public authorities** compiling inflation measures. It can certainly be an innovative source for the current production of official statistics, offering access to a wider set of prices (eg prices recorded online, credit card operations); it can also facilitate statistical compilation work when the data are easier to collect compared to traditional approaches. Indeed, data volumes have surged hand in hand with the development of specific techniques for their analysis, with the emergence of "big data analytics" – broadly referring to the general analysis of large data

---

[3] Following the work conducted under the aegis of the United Nations (see Meeting of the Expert Group on International Statistical Classifications (2015)).

[4] Note: internet-based indicators is not necessarily the category the most used in this context; see for instance the growing importance for public statisticians of the large datasets derived from administrative records (Bean (2016)).

sets – and "artificial intelligence" (AI); cf IFC (2019). Modern computing tools can now be used to collect data, correct them, improve coverage (eg web-scraping), process textual information (text-mining), match different data sources (eg fuzzy-matching), extract relevant information (eg machine learning) and communicate or display pertinent indicators (eg interactive dashboards). All these elements can help to address the resource issues posed by the compilation of official price statistics, especially in developing economies where statistical systems are still in infancy and staff skills are limited. One example is the Billion Prices Project at the Massachusetts Institute of Technology (MIT), which allows inflation indices to be constructed for countries that lack an official and/or comprehensive index and that can be used for enhancing international comparisons of price indexes in multiple countries and for dealing with measurement biases and distortions in international relative prices (see www.thebillionpricesproject.com and Cavallo and Rigobon (2016)). Similarly, a number of central banks in emerging market economies have compiled quick price estimates for selected goods and properties, by directly scraping the information displayed on the web, instead of setting up specific surveys that can be quite time- and resource-intensive. One notable situation relates to those developing economies as big as India, where collecting internet-based data is seen as a potentially useful alternative to the organisation of large surveys that would have to cover millions of reporters. Yet, as indeed noted by Hill (2018)[5] in the case of the United States, and in contrast to what is observed in the research and academic community, the use of big data in more mature statistical systems has been relatively incremental and limited. It is often targeted at methodological improvements (for instance quality adjustment) and at reducing reporting lags.

Turning to the measurement **challenges posed by rapid innovation**, the high velocity of big data sources can be particularly useful when prices change rapidly. For instance, direct web-scraping allows extracting almost in real time retailers' prices from online advertisements. This can support a timelier publication of official data, by bridging the time lags before official statistics become available – ie through the compilation of advance estimates or "nowcasting exercises". In addition to the lag issue, the information provided by the wide range of web and electronic devices is often available with a higher frequency; changes in price developments can thus be tracked more promptly, compared to official CPI numbers that are usually available on a monthly basis. This can be particularly useful when analysing early warning indicators and assessing turning points. Indeed, an important objective of the Billion Prices

[5] "… while nearly 15% of the price quotes in the Consumer Price Index are now collected online (…) the size of the CPI sample has not increased to reflect the lower cost of online data collection".

Project is not just to provide advance information on inflation, but also to estimate higherfrequency (daily) inflation indicators for a large number of countries, including advanced economies.

Third, the high granularity of big data sets can support to measure of **various dimensions of inflation,** allowing for a better understanding of the distribution of prices aggregates – eg across markets and/or locations. As regards market differentiation, web-scrapped prices can be useful for measuring inflation in very volatile sectors, such as fresh food prices, allowing for a better measure of some specific components of the CPI. As regards locations, large and granular data sets collected from commercial advertisements can help to capture local patterns with sufficient precision, say for instance to measure rents or property prices depending on zip codes or even street names.

Fourth, various big data sources such as numbers on internet search queries (eg Google Trends) and "soft" indicators computed from digitalised textual information (eg displayed by social media posts like Twitter) can provide interesting insights on **economic agents' sentiment and expectations** (Wibisono and Zulen (2019)). This can be particularly useful for nowcasting exercises and short-term forecasts (say, for the next quarters), as well as to assess the risks surrounding them. Traditional statistical surveys can also provide this kind of information, but they typically focus on specific items, eg firms' production expectations and consumer sentiment. In contrast, internet-based sources allow a wider range of indicators to be used. In addition, they can be less intrusive than face-to-face statistical surveys, and may therefore better reflect true behaviours and expectations (Rigobon (2018)).Yet, almost by construction, big data-based forecasts can mainly be used for **short-term forecasts**, since they depend on the flow of incoming data. In practice, longer-term inflation forecasts (say beyond the current and next year) have continued to remain model-based.

Fifth, new big data sources appear of increasing interest for **measuring the wide range of asset prices** that are not easily covered by traditional surveys because of the lack of statistics available and/or methodological clarity. Cases in point relate to residential and commercial property markets, which are often lacking reliable statistics, while alternative sources can be easily found using big data (eg advertisements from property websites and newspapers). In addition, these markets are characterised by a low and infrequent number of transactions (compared to stocks) and by significant heterogeneity across tangible assets, making the compilation of quality-adjusted house price indices difficult. These challenges can be overcome by capturing the various characteristics of the properties displayed in webbased advertisements and the application of hedonic methods. Moreover, the information collected, being very granular, can more easily be matched with other datasets, say census survey-based information for similar homes and tax registries. Furthermore, the new type of information collected can provide insights that are not covered by "traditional" statistics, eg to analyse housing market liquidity and tightness (by assessing demand intensity through the number of clicks on specific ads), discounting practices (by comparing asking and transactions prices, which can differ markedly for instance during turning points), and detailed geographical factors – see for instance Loberto et al (2018).

3. Challenges

Despite the various opportunities provided by big data sources, there are important challenges in using this information when measuring and forecasting prices. First, there are **practical difficulties** in collecting the data. This challenge can be reinforced by the large variety of big data formats, especially when the information collected is not well structured. Apart from the technical aspects (eg proper IT equipment, access rights etc), a key issue is data quality. For instance, references displayed on the web can be incorrect, or may not really reflect true transaction prices (eg in case customers benefit from discounts for other services, loyalty programs etc.); and the characteristics of the products may not be standardised properly. As a result, statisticians have to deal with duplicated information, since the same product may be sold in different places but is identified with different characteristics – for instance, a common feature for property markets is that several (different) advertisements can be associated with the same dwelling. Alternatively, a product may still be displayed on a website even though it is no more available for sale, hence the risk of measuring outdated prices. Dealing with these challenges requires significant work when cleaning and processing the data. In addition, the usefulness of the information collected is limited if the data sources and/or their market coverage change over time, and of course if its access is hindered by privacy laws and/or copyright issues.

There are also important **methodological limitations**. First, estimating price indices requires defining a basket of goods that are representative of the spending of the economic agents considered. As regards CPI, for instance, a significant part of the consumption basket is related to goods that are either not traded (eg self-consumption of housing services by homeowners) or that have an administrative nature and are therefore not quoted on the internet. So compiling a CPI indice using only web-based information will not be fully representative; one way to go is to complement this approach with other type of (non web-based) information.

Even if one only focusses on the part of the consumption basket that can indeed be traded on the internet, another concern is that **big data samples are often far from representative**, so the veracity of the information collected may not be as good as it seems. Certainly, big data sets usually cover entire populations, so by construction there is little sampling error to correct for. But a common misperception is that, because big data sets are extremely large, they are automatically representative of the true population of interest. Yet this is not guaranteed, and in fact the composition bias can be significant, in particular as compared with much smaller traditional probabilistic samples (Meng (2014)). For example, when measuring prices online, one must realise that not all transactions are conducted on the internet. The measurement bias can be problematic if online prices are different from the prices observed in physical stores, or if the products bought online are different to those sold in shops.

Lastly, there are also challenges when **using big data sources**. Ideally, statistics based on big data should have the same quality of standards and frameworks that govern official statistics,[6] such as transparency of sources, methodology, reliability and consistency over time. But in practice they can be collected in an opaque way, arguably not in line with these recognised principles. "Misusing" such information could thus raise ethical, reputational as well as efficiency issues. In particular, if the confidentiality of the data analysed is not carefully protected, this could undermine public confidence, in turn calling into question the authorities' competence in collecting, processing and disseminating information derived from big data (Tissot (2019)).

Using big data for **anticipating future developments** is also challenging. While related applications such as machine learning algorithms can excel in terms of predictive performance, they can lend themselves more to explaining what is happening rather then why. Indeed big data analytics rely frequently on correlation analysis, which can reflect coincidence as well as causality patterns. As such, they may be exposed to public criticism when insights gained in this way are used to produce official statistics and forecasts and/or justify policy decisions.

4. Conclusion

Big data sources and techniques can provide new and useful insights that **complement** "traditional" data sets and facilitate the compilation of price statistics as well as (short-term) forecasting exercises. They can also provide new sorts of signals that can be useful especially for policy makers, for instance for analysing market liquidity as well as geographical patterns.

[6] Cf the Fundamental Principles of Official Statistics adopted by the United Nations in 2014, available at unstats.un.org/unsd/dnss/gp/FP-New-E.pdf. For an overview of the challenges posed by using big data for official statistics more generally, see C Hammer et al (2017).

Central banks' experiences have underlined the benefits of having an **encompassing framework for making the most of all the data available** to enhance their regular monitoring of economic activity in general and of price developments in particular. These approaches draw on big data analytics to effectively summarise the information contained in large data sets through a small number of common factors and to capture time-varying effects so as to incorporate continuously the new information that becomes available (Bok and al (2017)). But, interestingly, they primarily rely on "traditional" survey statistics, while the use of web-based indicators has remained limited, suggesting that they may work less well in nowcasting/forecasting exercises compared to "traditional" statistics and confidence surveys. In the United States, for instance, the price nowcasting exercises conducted at the Federal Reserve Bank of Cleveland (Knotek and Zaman (2014)) mainly rely on past observations of inflation (eg monthly developments in CPI, core CPI, CPI for food, etc) complemented by higher-frequency information on oil prices (eg retail gasoline prices and daily crude prices); this reflects the fact that core inflation tends to be relatively stable in the short-run and that most of the volatility observed in headline inflation is driven by changes in food and energy prices. The parallel approach set up at the ECB relies on similar mechanisms when trying to predict short-term development in inflation (Modugno (2011): the high-frequency data retained for nowcasting inflation are mainly related to the daily prices of raw materials including energy. Such mixed approaches – eg using big data analytics to digest the large amount of incoming information while still having as input relatively "traditional" statistics – are reported to perform relatively well compared to, for instance, consensus forecasts. From a different perspective, the GDP nowcasting exercises conducted by the Federal Reserve Bank of New York have highlighted the information content of past inflation when estimating economic activity in advance.

Whether such approaches will allow for the **integration of more "big data-type" information when measuring and predicting inflation remains to be seen**. Certainly, there are important challenges to be considered, related in particular to: the methodological choices to be retained; the need to clean the vast of amount of the new information available, not least to deal with duplicates, outliers and other quality issues; the difficulty to measure real transaction prices and avoid capturing obsolete information (since internet-based prices can remain on the web for a long time); and the issues posed by matching different datasets at a quite granular level. Looking forward, key is to clarify/document the type of information available and allow researchers and policy analysers to test it in a transparent way.

References

1. Bean, C (2016): *Independent review of UK economic statistics,* March.
2. Bok, B, D Caratelli, D Giannone, A Sbordone and A Tambalotti (2017): "Macroeconomic nowcasting and forecasting with Big Data", *Federal Reserve Bank of New York Staff Reports*, no 830, November.
3. Bureau of Labor Statistics (BLS) (2019): *Handbook of Method*, Chapter 17. The Consumer Price Index, April.
4. Cavallo, A and R Rigobon (2016): "The Billion Prices Project: Using Online Prices for Measurement and Research", *Journal of Economic Perspectives,* Spring 2016, vol 30, no2, pp 151-78.
5. Croushore, D (1993): "Introducing: The Survey of Professional Forecasters", *Federal Reserve Bank of Philadelphia Business Review,* November/December.
6. Curtin R (1996): *Procedure to estimate price expectations*, University of Michigan Surveys of Consumers, January.
7. European Commission (2019): *The joint harmonised EU programme of business and consumer surveys - User Guide*, Directorate-General for Economic and Financial Affairs, January.
8. European Commission, IMF, OECD, United Nations and World Bank (2009): *System of National Accounts* 2008.
9. Garcia, J A (2003): "An introduction to the ECB's survey of professional forecasters", *ECB Occasional Paper Series*, no 8, September.
10. Groves, R (2011): "Designed data and organic data", in the Director's Blog of the US Census Bureau.
11. Hammer, C, D Kostroch, G Quirós and staff of the IMF Statistics Department (STA) Internal Group (2017): "Big data: potential, challenges, and statistical implications", *IMF Staff Discussion Notes*, no 17/06, September.
12. Hill, S (2018): "The Big Data Revolution in Economic Statistics: Waiting for Godot... and Government Funding", *Goldman Sachs US Economics Analyst,* 6 May.
13. ILO, IMF, OECD, UNECE, Eurostat and The World Bank (2004): *Consumer price index manual: Theory and practice*, Geneva, International Labour Office.
14. Irving Fisher Committee on Central Bank Statistics (IFC) (2017): "Big data", *IFC Bulletin*, no 44, September.
15. ——— (2019): "The use of big data analytics and artificial intelligence in central banking", *IFC Bulletin*, no 50, May.
16. Knotek E and S Zaman (2014): "Nowcasting US headline and core inflation", *Federal Reserve Bank of Cleveland Working Paper*, no 14-03, May.

17. Loberto M, A Luciani and M Pangallo (2018): "The potential of big housing data: an application to the Italian real-estate market", Bank of Italy *Temi* di *Discussione*, no 1171, April.
18. Meeting of the Expert Group on International Statistical Classifications (2015): *Classification of Types of Big Data*, United Nations Department of Economic and Social Affairs, ESA/STAT/AC.289/26, May.
19. Meng, X (2014): "A trio of inference problems that could win you a Nobel Prize in statistics (if you help fund it)", in X Lin, C Genest, D Banks, G Molenberghs, D Scott and J-L Wang (eds), *Past, present, and future of statistical science*, Chapman and Hall, pp 537–62.
20. Modugno, M (2011): "Nowcasting inflation using high frequency data", *ECB Working Paper Series*, no 1324, April.
21. Reinsdorf M and P Schreyer (2019): "Measuring consumer inflation in a digital economy", *OECD Statistics and Data Directorate Working Paper*, No 101, February.
22. Rigobon, R (2018): "Promise: measuring from inflation to discrimination", presentation given at the workshop on "*Big data for central bank policies*", Bank Indonesia, Bali, 23–25 July.
23. The Economist (2017): "The world's most valuable resource is no longer oil, but data", 6 May edition.
24. Tissot, B (2014): "Monitoring house prices from a financial stability perspective - the BIS experience", International Statistical Institute Regional Statistics Conference, November.
25. ——— (2018): "Providing comparable information to assess global financial stability risks", *Eurostat Statistical Reports*, KS-FT-18-001, January.
26. ——— (2019): "Making the most of big data for financial stability purposes", in S Strydom and M Strydom (eds), *Big Data Governance and Perspectives in Knowledge Management*, IGI Global, pp 1–24.
27. Wibisono O and A Zulen (2019): "Measuring stakeholders' expectations for the central bank's policy rate", *IFC Bulletin*, no 50, May.

Small area prediction of counts under nonparametric generalized linear mixed model

Hukum Chandra[1], Nicola Salvati[2], Ray Chambers[3]
[1] Indian Agricultural Statistics Research Institute, Library Avenue, India
[2] University of Pisa, Italy
[3] University of Wollongong, Wollongong, Australia

Abstract
We describe a methodology for small area estimation of counts that assumes an area-level version of a nonparametric generalized linear mixed model with a mean structure defined using spatial splines. The proposed method represents an alternative to other small area estimation methods based on area level spatial models that are designed for both spatially stationary and spatially non-stationary populations. We develop an estimator for the mean squared error of the proposed small area predictor as well as an approach for testing for the presence of spatial structure in the data and evaluate both the proposed small area predictor and its mean squared error estimator via simulations studies. Our empirical results show that when data are spatially non-stationary the proposed small area predictor outperforms other area level estimators in common use and that the proposed MSE estimator tracks the actual mean squared error reasonably well, with confidence intervals based on it achieving close to nominal coverage. An application to poverty estimation using household consumer expenditure survey data from 2011-12 collected by the national sample survey office of India is considered.

Keywords
Small area estimation; Nonparametric models; Spatial relationship; Count data; Poverty indicator

1. Introduction

When the variable of interest is binary or a count and small area estimates are required for these data, use of standard small area estimation (SAE) methods based on linear mixed models becomes problematic. For example, poverty indicators and many other indicators related to socio-economic status and food insecurity usually behave in a non-Gaussian manner at small area levels, and so estimation in these cases is typically based on a generalized linear mixed model (GLMM); see Manteiga et al. (2007) and Ruppert et al. (2003, chapter 10). In many applications this is not possible, for example poverty mapping where data confidentiality restricts access to unit level survey data with small area identifiers, or where the agency carrying out the small area analysis does not have the resources to analyse unit level data, as in many developing countries. In such situations, an area level version of the GLMM can be used for SAE. In particular, when only area level data are available, an area level version of the GLMM is fitted to obtain the plug-in empirical predictor for the small areas, see for example, Chandra et al. (2011). In economic, environmental and epidemiological applications, estimates for areas that are spatially close may be more alike than estimates for areas that are further apart. It is therefore reasonable to assume that the effects of neighbouring areas, defined via a contiguity criterion, are correlated. Chandra and Salvati (2018) describe an extension of the area level version of GLMM that allows for spatially correlated random effects using a SAR model (SGLMM) and define a plug-in empirical predictor (SEP) for the small area proportion under this model. This model allows for spatial correlation in the error structure, while keeping the fixed effects parameters spatially invariant. Chandra et al. (2017) introduce a spatially nonstationary extension of the area level version of GLMM, using an adaptation of the geographical weighted regression (GWR) concept to extend the GLMM to incorporate spatial nonstationarity (NSGLMM), which they then apply to the SAE problem to define a plug-in empirical predictor (NSEP) for small areas. Non-stationary spatial effects can be also modelled using a spatially non-linear extension of the GLMM. In the GLMM, the relationship between the link function and the covariates is often assumed to be linear. However, when the functional form of the relationship between the link function and the covariates is unknown or has a complicated functional form, an approach based on the use of a non-linear regression model can offer significant advantages compared with one based on a linear model. When geographically referenced area-level responses play a central role in the analysis and need to be converted to maps, we can use bivariate smoothing to fit a spatially heterogeneous GLMM. In particular, we use P-splines that rely on a set of bivariate basis functions to handle the spatial structures in the data, while at the same time including small area random effects in the model. We denote this nonparametric P-spline-based extension of the usual GLMM by SNLGLMM. See Opsomer et al. (2008) and Ruppert et al. (2003). We then describe a non-linear version of the plug-in empirical predictor for small areas (SNLEP) under an area level version of SNLGLMM. We also develop mean squared error estimation for SNLEP using the approach discussed in Chandra et al. (2011) and Opsomer et al. (2008).

2. SAE under a Spatially non-linear generalized mixed model

Consider a finite population U of size N, and assume that a sample s of size n is drawn from this population according to a given sampling design, with the subscripts s and r used to denote quantities related to the sampled and non-sampled parts of the population. We assume that population is made up of m small domains or small areas (or domains or areas) $U_i (i=1,...,m)$, where we use a subscript of i to index those quantities associated with area i. In particular, n_i and N_i are used to represent the sample and population sizes in area i, respectively. We also assume that the underlying unit level variable of interest y is discrete, and in particular is either a binary value or is a non-negative integer, and the aim is to estimate the corresponding small area population proportions or population totals (i.e. counts). Let the total of y in area i be denoted y_i, and let y_{si} and y_{ri} denote the corresponding sample and non-sample counts for area i respectively. We shall assume that area level auxiliary information from secondary data sources, e.g., Census and Administrative records, is available. Let x_i be the p-vector of these covariates for area i from these sources. The area level version of the GLMM is then defined as Pr($y_i \mid x_i$) $\propto \pi_i$, where

$$g(\pi_i) = \eta_i = X_i^T\beta + u_i \tag{1}$$

where g(·) is a known function, called the link function, $\pi_i = g^{-1}(\eta_i)$, β is the p-vector of regression coefficients, often referred to as the fixed effect parameter of the GLMM, and $u_i \sim N(0, \sigma_u^2)$. The model (1) can be used to relate the area level direct survey estimates to area level covariates. This type of model is often referred to as 'area-level' model in SAE (Fay and Herriot,1979). Collecting the area level models (1), we can write the model (1) as

$$g(\pi) = \eta = X\beta + u, \tag{2}$$

where $\pi = (\pi_1, \ldots\ldots, \pi_m)^T$, $X = (x_1^T, \ldots\ldots, x_m^T)^T$ is a $m \times p$ matric and u = $(u_1, \ldots\ldots, u_m)^T$ is a vector of $m \times 1$ of area random effects which is normally distributed with mean zero and variance $\sum_u = \sigma_u^2 I_m$. Here, I_m is an identity matrix of order m. When the variable of interest y is binary, and unit level values in area i are independently and identically distributed, the sample counts y_{si} in area i, has a Binomial distribution with parameters n_i and π_i , denoted by y_{si} ~ Binomial(n_i, π_i), where π_i is now the probability of occurrence of an event or probability of prevalence in area i, often referred to as the probability of a 'success'. Similarly, the non-sample count y_{ri} in area i is such that y_{ri} ~ Binomial(N_i - n_i, π_i). That is, the counts y_{si} and y_{ri} are independent Binomial variables with π_i then corresponding to a common success probability. In this case, the link function g(·) is usually taken to be the logit of the probability π_i. The model (1) linking π_i with the covariates x_i is then the GLMM with logistic link function given by logit(π_i)= $\ln\{\pi_i(1-\pi_i)^{-1}\} = \eta_i = x_i^T\beta + u_i$, with $\pi_i = \exp(\eta_i)\{1 + exp(\eta_i)\}^{-1} = expit(\eta_i) = expit(x_i^T\beta + u_i)$ and $u_i \sim N(0, \sigma_u^2)$. Here, $y_{is} \mid u_i \sim \text{Binomial}\left(n_i, expit(x_i^T\beta + u_i)\right)$ and $y_{ir} \mid u_i \sim \text{Binomial}\left(N_i - n_i, \text{expit}(x_i^T\beta + u_i)\right)$. The expected values of y_{si} and y_{ri} given u_i are then $\mu_{si} = E(y_{si}|u_i) = n_i \,\text{expit}(x_i^T\beta + u_i)$ and $\mu_{ri} = E(y_{ri}|u_i) = (N_i - n_i)\text{expit}(x_i^T\beta + u_i)$. The population count in area i can be expressed as $y_i = y_{si} + y_{ri}$, where the first term y_{si}, the sample count, is known whereas the second term y_{ri}, the non-sample count, is unknown. A plug-in empirical predictor (EP) of the population count in area i is obtained by replacing y_{ri}, by its predicted value $\hat{\mu}_{ri} = \hat{E}\ (y_{ri}|u_i)$ under the model (1) as

$$\hat{y}_i^{EP} = y_{si} + \hat{\mu}_{ri} = y_{si} + (N_i - n_i)\hat{\pi}_i^{EP}, \tag{3}$$

with $\hat{\pi}_i^{EP} = \text{expit}\left(x_i^T\hat{\beta} + q_i^T\hat{u}\right)$ for binary data, where $q_i^T = (0, \dots, 1, \dots, 0)$ is $1{\times}m$ vector with 1 in the i-th position and $\hat{u} = (\hat{u}_1, \dots, \hat{u}_m)^T$. In SAE problems, the sample size n_i is often negligible relative to the population size N_i, then $\hat{y}_i^{EP} = N_i\hat{\pi}_i^{EP}$. An estimate of the proportion in area i is given by $\hat{\pi}_i^{EP}$.

We now introduce a spatially non-linear extension of an area level GLMM. We refer this model as spatially non-linear generalized linear mixed model (SNLGLMM). We start by developing the nonparametric extension of the GLMM, and then suggest a spatial extension of this model. Typically, the fixed effect part of a GLMM is assumed to be linear. However, in reality the functional form of this relationship may be unknown or it may have a complicated functional form. Without loss of generality we restrict our development to the case of a single covariate x and use nonparametric regression modelling based on a P-spline approximation. The spatially non-linear GLMM (SNLGLMM) is then of the form

$$g(\pi_i) = \eta_i = f(x_i) + u_i \tag{4}$$

where $u_i \sim N(0,\ \sigma_u^2)$ is the area specific random effect and $\pi_i = E(y_i|u_i, x_i) = h\{f(x_i) + u_i\}$. In particular, the spatially non-linear logistic-normal mixed model and the spatially non-linear Poissonnormal mixed model for binary and count data, respectively, are defined as $logit(\pi_i) = \eta_i = f(x_i) + u_i$ with $\pi_i = expit\,\{f(x_i) + u_i\}$ and $\log(\pi_i) = \eta_i = f(x_i) + u_i$ with $\pi_i = exp\{f(x_i)\}$. The function $f(x_i)$ in (4) is unknown, but can be approximated sufficiently well by the P-spline approximation

$$f(x_i, \beta, \gamma) = \beta_0 + \beta_1 x_i + \cdots \dots + \beta_p x_i^p + \sum_{l=1}^{L} \gamma_l (x_i - K_l)_+^p \tag{5}$$

Here p is the degree of the spline, $(t)_+^P = t^P$ if $t > 0$ and its otherwise, K_l for l= 1, ……, L is a set of fixed constants called knots, $\beta = \left(\beta_0, \beta_, \dots, \beta_p\right)^T$ is the coefficient vector of the parametric portion of the model and $\gamma = (\gamma_1, \dots, \gamma_L)^T$ is the vector of spline coefficients, L is the number of spline knots, and $\gamma_l \sim N\left(0, \sigma_\gamma^2\right)$; l = 1,……, L. Provided that the knot locations are sufficiently spread out over the range of x and L is sufficiently large, the class of functions defined by (5) can approximate most smooth functions. Ruppert et al. (2003, chapter 5) suggest the use of a knot for every four observations, up to a maximum of about 40 knots for a univariate application. This is usually done by placing these knots at equally spaced quantiles of the distribution of the covariate.

Note that the P-spline approximation consists of a linear combination of appropriately chosen basis functions. For simplicity, the approximating function $f(x,\beta,\gamma)$ in (5) uses truncated polynomial spline basis functions. $\{1, x, \ldots, x^p, (x-K_1)_+^p, \ldots, (x-K_L)_+^p\}$. Other basis functions, e.g. B-splines (Eilers and Marx, 1996) or radial functions, can also be used. Using a large number of knots in expression (5) can lead to an unstable fit. In order to overcome this problem, a penalty is usually put on the magnitude of the spline parameters γ.

When geographically referenced responses play a central role in the analysis and need to be converted to maps, we can use bivariate smoothing, $f(x_{1i}, x_{2i}) = f(x_{1i}, x_{2i}, \beta, \gamma)$ where x_{1i} and x_{2i} are spatial coordinates. This is usually the case with environmental, agricultural, public health and poverty mapping applications. Consequently we assume the following model (Opsomer et al., 2008)

$$f(x_{1i}, x_{2i}, \beta, \gamma) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + z_i\gamma, \tag{6}$$

where z_i is the i-th row of the following $n \times L$ matrix

$$Z = [C(x_i - K_l)]_{\substack{1<i<n \\ 1<l<L}} [C(K_l - K_{l'})]_{1<l<L'}^{-1/2} \tag{7}$$

Where $C(t) = \|t\|^2 \log\|t\|$, $x_i = (x_{1i}, x_{2i})$ and $K_l (l = 1, \ldots, L)$ are knots. Note that the C(t) function is defined so that when there is a knot at every observation (that is, the full rank case) this model for bivariate smoothing leads to a thin plate spline (Green and Silverman, 1994). The second matrix on the right-hand side of (7) applies a linear transformation to the radial basis functions defining the first matrix, and was recommended by Ruppert *et al.* (2003) as a way of making the radial spline behave approximately like a thin plate spline. More details on the Z matrix can be found in Ruppert *et al.* (2003, chapter 13). As suggested by Ruppert *et al.* (2003), fitting the approximation (5) can be simplified by treating the vector γ as a random-effect vector in a mixed model specification, which allows estimation of β and γ by maximum likelihood methods. Following Opsomer *et al.* (2008), Wand (2003) and Ruppert *et al.* (2003, chapter 4), the spline approximation to (4), and to (6), can be written as

$$g(\pi) = \eta = X\beta + Z\gamma + u. \tag{8}$$

If other covariates are available, they can be included in the model as parametric terms, by being added to the matrix X. In this model γ is assumed to be a Gaussian random vector of dimension L. In particular, it is assumed that $\gamma \sim N_L(0, \sum_\gamma = \sigma_\gamma^2 I_L)$, where I_t denotes the identity matrix of dimension t. Here u is the m-vector of random area effects. As usual, it is assumed that the area effects in u are distributed independently of the spline effects in γ with $u \sim N_m(0, \sum_u = \sigma_u^2 I_m)$. Under (8), a plug-in spatially non-linear empirical predictor (SNLEP) for the total count y_i in area i is given by

$$\hat{y}_i^{SNLEP} = y_{si} + (N_i - n_i)\hat{\pi}_i^{SNLEP}, \tag{9}$$

Where $\hat{\pi}_i^{SNLEP} = \text{expit}(x_i^T\hat{\beta} + z_i^T\hat{\gamma} + q_i^T\hat{u})$ for binary data and $\hat{\pi}_i^{SNLEP} = \exp(x_i^T\hat{\beta} + z_i^T\hat{\gamma} + q_i^T\hat{u})$ for count data, and x_i^T, z_i^T and q_i^T donate respectively the rows of X, Z and $\mathrm{I_m}$ that correspond to area i. When n_i is negligible compared to N_i, the SNLEP (9) is $\hat{y}_i^{SNLEP} = N_i \hat{\pi}_i^{SNLEP}$. An estimate of the proportion or rate in area i is $\hat{\pi}_i^{SNLEP}$. In the spirit of Chandra *et al.* (2017) and Opsomer *et al.* (2008), we develop a bootstrap procedure to test the spatial nonlinearity hypothesis, that is, the hypothesis $H_{0\sigma_\gamma^2}$: $\sigma_\gamma^2 = 0$ versus the one-sided alternative $H_{1\sigma_\gamma^2}$: $\sigma_\gamma^2 > 0$.

3. Empirical Evaluations

This Section presents the results from simulation studies that compare the empirical performance of the proposed SNLEP estimator (9) under the SNLGLMM (4) with the EP (3) under the GLMM (1). The performance criteria used are the percentage Relative Bias (RB) and the percentage Relative Root MSE (RRMSE). In our model based simulations we set the number of small areas m = 100 and considered two values for the area specific sample sizes n_i=10 and 50 with $N_i = 100$ and 5000, respectively.

We used an area level version GLMM to generate data. The response values were generated from $y_i \sim \text{Binomial}\,(n_i, \pi_i)$ and $\text{logit}(\pi_i) = \eta_i = f(x_i) + u_i$ with $\pi_i = \exp(\eta_i)\{1 + \exp(\eta_i\}^{-1}$. Spatial locations were simulated as the values of two independently distributed uniform [0,1] covariates x_1 and x_2, and the random area effects u_i were generated as m independent realizations from a $N(0, \sigma_u^2 = 0.0625)$ distribution. We considered two different choices of the response function $f(x_1, x_2)$: Plane $f(x_1, x_2) = 0.5x_1 + 0.2x_2$,

$$\text{Mountain : } f(x_1, x_2) = \frac{40\exp[8\{(x_1-0.5)^2 + (x_2-0.5)^2\}]}{\exp[8\{(x_1-0.2)^2+(x_2-0.7)^2\}]+\exp[8\{(x_1-0.7)^2+(x_2-0.2)^2\}]}$$

A total of T = 1000 data sets were independently generated under each of these models and the predicted small area counts for the two predictors developed in the previous Sections were calculated. Note that in these simulations the SNLGLMM uses the radial basis functions with L=25 knots, following Pratesi *et al.* (2009).

Table 1 shows the average values of percentage relative bias (RB) and the average values of percentage relative root MSE (RRMSE) recorded by the SAE methods investigated in our simulations. In particular, Table 1 sets out the average relative biases and the average relative RMSEs of the two small area predictors (EP and SNLEP) across the two different types of response function $f\,(x_1, x_2)$ (i.e. plane and mountain) used in SNLGLMM for data generation, allowing one to compare the two predictors across different data types. The results in Table 1 are essentially as one would expect. Here we clearly see that the performances of EP and SNLEP are on a par when data are generated using the plane function. In contrast, the simplicity of the EP predictor comes at a price when data are generated using the mountain response function. In this case, the SNLEP performs best, with low bias and is more efficient than the EP. As expected, these results improve as the sample size increases.

Table 1. Percentage relative bias (RB) and the percentage relative root MSE (RRMSE) of the EP, and SNLEP with $n_i = 10$ and $n_i = 50$ under the plane and the mountain response functions. Values are averages over the small areas.

f (*x*)	Predictors	$n_i = 10$		$n_i = 50$	
		RB	RRMSE	RB	RRMSE
		Binary data			
Plane	EP	-0.073	12.89	-0.029	9.51
	SNLEP	-0.074	13.95	-0.029	9.80
Mountain	EP	6.320	26.65	1.729	15.96
	SNLEP	1.587	20.61	0.527	13.22

4. Concluding remarks

This paper describes a spatially non-linear (or nonparametric) extension of the area level version of the generalized linear mixed model (SNLGLMM) and considers SAE under this model. The corresponding estimator is referred to as the spatially non-linear empirical predictor (SNLEP) for small areas. This estimator can accommodate situations where the functional form of the spatial relationship between the variable of interest and the covariates is unknown. Empirical evaluations based on simulation studies indicate that the proposed SNLEP method is more efficient than the EP under the area level generalized linear mixed model. Although details are not reported here but the proposed analytic MSE estimator also performed reasonably well, with good coverage performance for nominal confidence intervals based on it. We also applied the SNLEP to real survey data to estimate the Head Count Rate (HCR) poverty indicator values for the districts of the State of Uttar Pradesh in India and produced a poverty map of these districts based on these HCR estimates.

References

1. Chandra, H. and Salvati, N., 2018. Small area estimation for count data under a spatial dependent aggregated level random effects model. Communications in Statistics - Theory and Methods, 47 (5), 1234 -1255.
2. Chandra, H., Salvati, N. and Chambers, R., 2017. Small area prediction of counts under a nonstationary spatial model. Spatial Statistics 20, 30-56.
3. Chandra, H., Salvati, N. and Sud, U.C., 2011. Disaggregate-level estimates of indebtedness in the state of Uttar Pradesh in India-an application of

small area estimation technique. Journal of Applied Statistics 38(11), 2413-2432.

4. Eilers, P. H. C. and B. D. Marx, 1996. Flexible smoothing with B-splines and penalties. Statistical Science 11 (2), 89–121.
5. Fay, R. E. and Herriot, R. A., 1979. Estimation of income from small places: an application of James-Stein procedures to census data. Journal of the American Statistical Association 74, 269277.
6. Green, P.J. and Silverman, B.W., 1994. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Chapman & Hall Ltd.
7. Manteiga, G.W., Lombardìa, M.J., Molina, I., Morales, D., and Santamarìa,L., 2007. Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model. Computational Statistics and Data Analysis 51, 2720-2733.
8. Opsomer, J.D., Claeskens, G., Ranalli, M.G., Kauermann, G. and Breidt, F.J., 2008. Nonparametric small area estimation using penalized spline regression. Journal of the Royal Statistical Society B 70, 265-286.
9. Pratesi, M., Ranalli, M.G., Salvati, N., 2009. Nonparametric M-quantile regression using penalised splines. Journal of Nonparametric Statistics 21, 287-304.
10. Ruppert D., Wand M.P., Carroll R.J., 2003. Semiparametric Regression. Cambridge University Press, New York.

A case study of data integration in Australia

Michael Beahan
Australian Bureau of Statistics

Abstract

The Australian Bureau of Statistics (ABS) is collaborating with five other Australian Commonwealth agencies to produce high-quality and secure integrated data. In this paper, the Multi-Agency Data Integration Project (MADIP) will be highlighted. MADIP contains information on healthcare, education, government payments, personal income tax, and the Census. These data are being combined to create a comprehensive picture of Australia. This project demonstrates the policy potential when existing public data are combined. The specific case of targeted services (such as healthcare) to communities in need is discussed.

Keywords

Data Integration; partnerships; policy insights

Data Integration – The Broader Context

Historically, the primary data sources at the Australian Bureau of Statistics (ABS) revolved around the conduct of surveys and censuses. Generally, the role of administrative data was secondary –as a comparison to survey and census data and a means to test its correctness.

Alternative data sources (to surveys and censuses conducted by a national statistical office) come in many shapes and sizes; and from both government and private sectors. In the past (and to a lesser extent today) the utility of data sourced from other government agencies for the purposes of official statistics, policy formulation and evaluation has been limited. Long standing data sharing arrangements across the Australia public service are complex and typically hinder the efficient and effective use of data.

Barriers to greater sharing of data within government include:

- a dense web of legislative requirements which lack consistency;
- a culture of risk aversion, leading to overly cautious legislative interpretation and approval process complexity; and
- the lack of a whole-of-government approach.

In Australia in more recent times there has been a concerted effort to maximise the entirety of data holdings across government. The Australian Government recognises that the data it holds is a strategic national resource with potential to provide transforming policy outcomes for all Australians.

Greater use and sharing of public data facilitates opportunities for enhanced productivity, increased competition, improved delivery of government services and research outcomes.

Many high priority public policy challenges, such as homelessness, climate change, and crime, do not fit neatly within one ministerial portfolio or a single agency's set of responsibilities. Similarly, the information needed to address complex policy issues and research questions are often spread across a number of agencies.

Analysis of integrated datasets offers valuable opportunities to investigate more complex and expanded policy and research questions than would be possible using only separate, unlinked data sources. Data integration can produce new official statistics to inform society.

Data integration can reduce the need for costly collections by better leveraging existing data to meet current and emerging information requirements. Maximising the use of existing data, rather than establishing new collections, avoids additional load on respondents, helps to ensure cost-effectiveness and can improve timeliness. Data integration is therefore a key strategy for maximising governments' investments in existing information assets.

An exemplar of this new approach to data sharing is the Multi-Agency Data Integration Project.

Multi-Agency Data Integration Project (MADIP)

The Multi-Agency Data Integration Project (MADIP) is a partnership among Australian Government agencies to combine information on healthcare, education, government payments, personal income tax, and the Census to create a comprehensive picture of Australia over time.

Authorised researchers can use unidentified MADIP data to look at patterns and trends in the Australian population; and provide insight into the effectiveness of government policies, programs, and services. MADIP demonstrates how combining existing public data can help target services (such as healthcare) to the people and communities who need them.

MADIP is a core component of the Australian Government's Data Integration Partnership for Australia (DIPA). There are six Commonwealth agencies working together on the project:

- Australian Bureau of Statistics
- Australian Taxation Office
- Department of Education and Training
- Department of Health
- Department of Human Services
- Department of Social Services.

MADIP is an information asset still in its formative years yet already this unique combination of previously disparate data sources is providing unique insights.

Cases Studies

<u>People living in remote areas are less than half as likely to access a mental health service</u>

Policies and services can only be as effective as the information underpinning them, and it was identified that there was a critical gap in Australia's understanding of people accessing mental health-related services and prescriptions. Using data integration, it has been possible to fill in that gap without needing to carry out another round of surveys.

The Mental Health Services and Census project integrates public health data and census data, and has contributed significantly to the pool of mental health-related research data in Australia. This data is being used in the development and evaluation of mental health programs and support services now and into the future. Questions can be answered about people accessing subsidised mental health-related services and medications with evidence that up until now has not been available. For example, the enhanced Census data was critical in informing the 2014 National Review of Mental Health Programmes and Services Report and helped join the dots between mental health-related services, medication use, and key demographic information such as education, employment and housing.

One important insight from the data suggested that when accessing mental health services, people with less education, the unemployed, and those living in rural areas are more likely to be prescribed drugs, while more educated and city-based patients were more likely to be prescribed talking-based therapies. This data enables government to understand the inconsistency, and create policies to address it.

For Dr Paul Jelfs, General Manager of the Population & Social Statistics Division at the Australian Bureau of Statistics, the project has improved our capacity to support people with mental health issues. Quoting Paul, "If you're running a mental health service, your service will be more effective if you understand the employment and education characteristics of patients in your area."

In this study, 2011 Census data was enhanced with the reuse of administrative information from the Medicare Benefits Schedule (MBS) and the Pharmaceutical Benefits Scheme (PBS). The National Mental Health Commission approached the ABS to integrate this data, and the data was processed in a confidential secure facility, with personal privacy and confidentiality guaranteed by law.

Outcomes for permanent migrants in the Australian labour market

The integration of Australian Government datasets is providing new insights into the migrant experience and helping to inform government immigration policy.

The Australian Productivity Commission used two integrated datasets during its 2015 public inquiry on Australia's migrant intake to conduct deeper and more definitive analysis into the variability of migrant outcomes in the Australian labour market. These were the Australian Census and Migrants Integrated Dataset (ACMID) and the Personal Income Tax and Migrants Integrated Dataset (PITMID).

These datasets were developed by integrating Australian Bureau of Statistics (ABS) Census data and Australian Taxation Office (ATO) data with the Australian Government's Settlement Database. This integration work was done collaboratively by the ABS, the ATO, the Department of Social Services (DSS) and the Department of Immigration and Border Protection (DIBP).

One focus of the inquiry was the effects of substantially increasing the charges associated with permanent migrant visas. The Commission developed a model, drawing on the ACMID and other supplementary datasets, to explore the impacts of increased use of immigration charges on the composition of Australia's permanent migrant intake, including on fiscal outcomes. This modelling enabled the Commission to make policy recommendations based on evidence.

The Productivity Commission's 2016 Inquiry Report Migrant Intake into Australia found that there were substantial differences for permanent migrants in the Australian labour market depending on their visa stream.

Aboriginal and Torres Strait Islander Life Expectancy

When the Council of Australian Governments (COAG) committed to closing the gap in life expectancy between Aboriginal and Torres Strait Islander and non-Indigenous Australians by 2031, one of the initial challenges was data quality and reliability. Data suggested that the gap in child mortality had reduced, while the life expectancy gap was showing little improvement. But before you can be confident of how to fix something this important and multi-faceted, you need a detailed understanding of the problem.

To help with this, the ABS Mortality Project matched existing death registrations with data from the 2011 Census, to enable a clearer picture of life expectancy for Aboriginal and Torres Strait Islander peoples. The enhanced data provides a benchmark by which COAG goals are measured, and ultimately measures the success of closing the gap for life expectancy. Policy makers can use findings from this data as they chart the course to policy change and better outcomes.

As an Aboriginal leader, public health policy advisor and health data user, University of Canberra Chancellor Professor Tom Calma AO said policy makers and governments need the most accurate data possible to inform good decision making. Professor Calma made the following points:-

- "While no health intervention will be effective unless the community own it and drive it, more detailed and more accurate data is important, particularly when we consider social and cultural determinants of health,"
- "Mortality data is one of the real statistical measures of a population's long term health and wellbeing - you've got to recognise these statistics are people, they're real.
- "When we look at Infant Mortality Rates it's extreme for Indigenous people. We want to see that there are improvements and try to identify the catalysts that have led to improvements."
- "Without the ability to pull together data sets we would be in a 'hit and miss' situation and we ignore the data at our peril. We need to move away from politicians and their gut feelings to looking at the data and what it is telling us and using it to drive the interventions."

The Australian Institute of Health and Welfare (AIHW) uses ABS mortality data to inform the Australian Burden of Disease Study, exploring fatal disease among Aboriginal and Torres Strait Islander peoples. The ABS' data integration work on this project and others provides policy makers and the community with more accurate data while maintaining the highest level of individual privacy for respondents.

The ABS Mortality Project is giving policymakers a better measure of Aboriginal and Torres Strait Islander life expectancy in Australia. This improved data accuracy means the COAG goal of closing the mortality gap between Aboriginal and Torres Strait Islander peoples and non-Indigenous Australians can be better informed and monitored.

<u>Education</u>

What difference does in-school vocational education and training (VET) make to Year 10-12 students? Does VET in school improve Year 12 retention? How likely are those students to go on to further education? What are their employment outcomes? How can we improve these outcomes?

In answering those questions, the Census and Vocational Education and Training in Schools project addressed a critical gap in our understanding of the post-school outcomes for students undertaking VET in Schools.

Where policy makers used to make assumptions about efficacy and impact, they now have evidence. In fact, the project demonstrated that students who do VET in Schools and do not go on to higher education have better engagement and employment outcomes. It also showed that male

Aboriginal and Torres Strait Islander students who do VET in Schools are more likely to complete Year 12 and that male VET in Schools students who study a trade gain better employment outcomes.

Dr Patrick Korbel from the National Centre for Vocational Education Research (NCVER) said that while the annual VET in Schools data gave information about students in the year of collection, it didn’t provide any post-school insights.

Quoting Dr Korbel:-

- “However, by linking NCVER’s training data from 2006 with 2011 Census data we could see the educational and employment outcomes of those students 5 years down the track,”
- “We could see whether the students went on to any further study, the highest level of qualification they attained, whether they were currently employed or still studying and the level of income they had attained. All of this information could paint the picture of the benefits of the VET training they had undertaken at school.”
- “By linking NCVER’s VET data about students’ education with Census employment outcomes we capture new and important information. It’s really valuable information and it revealed that one of the main predicted benefits of in-school vocational training – to keep students in the school system until they finish Year 12 – is happening,”
- “The new standard is a Year 12 education – gone are the days of leaving school at Year 10 and taking up a trade. Now we can show who is participating in the workforce and the outcomes they are achieving. It’s a great basis for future research to see what is working and what is not,”
- “Surveys are expensive and time consuming and it’s very difficult to trace these students once they leave school. If you wanted to follow up with another survey it would not only be cost prohibitive, it would be unfeasible to track them,”
- “Data integration is crucial and the premise is collect once, use many times. Now that we have linked this data once, it will be easier next time and since NCVER collects information from across the VET sector, the opportunities are endless to learn about the outcomes these programs and policies achieve.”

The Road Forward

Of course none of this is easy and certainly doesn’t happen overnight. The ABS has had to navigate a range of issues in arriving at a point where we now are significantly better able to provide a service to the Australian government and its associated research communities.

In 2018 the Australian Government established the Office of the National Data Commissioner. The Office of the National Data Commissioner will be responsible for implementing a simpler data sharing and release framework. The new framework will break down the barriers which prevent efficient use and reuse of public data, while maintaining the strong security and privacy protections that the community expects. A central pillar of this new framework will be new Data Sharing and Release Legislation

Community attitudes and expectations are central to the ongoing success of data integration projects and, in general, communities are more accepting of such projects on the back of two key elements; maintenance of individual privacy and demonstrated benefits.

The success of projects such as MADIP has placed a significant demand on the ABS in terms of the availability of data. As the numbers of researchers, projects and datasets grows exponentially the ABS has been considering cloud based technology as the best means of supplying scalable IT solutions.

The growth of Data Integration projects within government has also highlighted the growing need for data analytic skills both within ABS and across government. Data skills and capability are as critically important for the Australian Public Service (APS) as anywhere else. Data literacy across the government have a critical role in supporting evidence-based decision making, developing more efficient government policy and delivering services that meet the needs of people across Australia. Skills and knowledge in publishing, linking and sharing public data will help to make government services more citizen-focused.

Data remains the essential ingredient. However, the 21st century is emerging as a very dynamic environment for statistical agencies well away from our traditional partnerships, tools, methods and processes.

The next generation of International Statistical Standards

Gabriel Quirós-Romero, Jennifer Ribarsky
International Monetary Fund. GQuiros@imf.org and JRibarsky@IMF.org. The views expressed are those of the authors and do not necessarily represent the views of the International Monetary Fund (IMF), its Executive Board, or IMF management.

Abstract
Advances in technology and communication, increasing capital movements and dominance of multinationals as well as reductions in shipping costs have redefined the global economy. Leading to tighter integration of economies, with the creation of global value chains through which multinational corporations stretch their operations across borders. Current statistical manuals, for which value-added generation is anchored in the traditional notion of residency, are showing their limits.
Digitalization is also changing how we work, play and communicate in manners that stretch our statistical standards. Likewise, the gig economy, facilitated by digitalization, has given a new meaning to the concept of informality, but may help better measure these activities using big data. Furthermore, data is "the new oil" of modern economies, a valuable corporate asset. Lastly, distributional issues are poorly accounted for in the SNA whose metrics depicts aggregates or averages. This paper offers a discussion on these issues.

Keywords
SNA; Balance of payments; Globalization; Digitalization; Informal economy

1. Introduction
The International Monetary Fund's (IMF, the Fund) primary purpose is to ensure global stability of the international monetary system which includes all macroeconomic and financial sector issues that bear on global stability. The Fund does this in three ways: economic surveillance[1], lending, and capacity development. Statistics are an important component for ensuring sound policy analysis and appropriate evidence-based policy responses.

The international standards explain concepts, definitions, classifications, and accounting rules that comprise the internationally agreed standards for measuring the relevant concept (e.g., gross domestic product (GDP)) and are designed to be robust and flexible so that needs of the various users are met.

[1] The provision of economic and financial information needed for surveillance is spelled out in the Articles of Agreement of which Article VIII, Section 5 is a central pillar.

It has been over ten years since the adoption of the 2008 System of National Accounts (SNA) and the IMF's 6th edition Balance of Payments and International Investment Position Manual (BPM6) and economic systems have continued to evolve. The IMF's Government Finance Statistics Manual (GFS 2014) and the Monetary and Financial Statistics Manual and Compilation Guide (MFS 2016) were updated more recently. There are good reasons for maintaining separate manuals across specific topical domains (e.g. MFS meets the data needs for monetary statistics) but demands from the user community to have consistent statistics across domains and statistical compilers to have consistent guidance point towards a harmonized conceptual approach to standard setting. Thus, an overarching goal is to maintain the theoretical linkages across macroeconomic statistics while maintaining separate manuals.

While the current conceptual frameworks of national accounts (SNA) and international accounts (BPM6) are harmonized, robust, resilient and still relevant, it is also clear that globalization, digitalization, economic welfare, "gig" employment, or the informal economy provide significant challenges, especially in the case where the issues become entangled. Concerns have arisen over how these issues are being addressed both from a conceptual and a practical measurement perspective.

The IMF Statistics Department's (STA) mandate is to provide global leadership on statistical methodologies and standards for the Fund, its member countries, and the international statistical community, as well as strengthen member countries' capabilities to produce and disseminate macroeconomic and financial statistics through technical assistance and capacity development. STA's challenge is to provide guidance to a wide range of countries with very different economic structures and statistical systems, making the aspiration of one-statistical-standard-fits-all particularly difficult to implement. Therefore, a balance must be struck between methodology and statistical capacity in the IMF membership especially for less statistically advanced countries. This paper is organized by first providing some necessary reflections to consider before changing the international standards, then discusses main priority issues and recent progress.

2. Some initial reflections

The national and international accounts are a system of accounts that are interlinked, comprehensive, consistent, and integrated. Changes to one component may have implications that ripple throughout the fundamental principles, e.g. residence (territory), production boundary, asset boundary, as well as the "core", of the entire system of accounts. For an orderly discussion, it is convenient to sort issues by the complexity of the potential solutions. Hence, we propose first to reflect on whether the issue can be remedied by (i) just clarifying the recording of the new phenomena and/or providing more detailed guidance, without affecting the fundamental framework; (ii) providing (more) disaggregated (i.e. more granular) statistics, offering institutional sector, firm characteristics (e.g., ownership, industry, and size) or distributional information; (iii) rearrangement of classifications to reflect new agents (e.g., digital platforms), transactions, goods and/or services, however, all of them supplementary to related statistics; (iv) changing the fundamentals of the international standards.

Such distinctions are key when considering when an update of the manuals is warranted, and if so with which depth and range. Issues that do not change the fundamentals of the international standards could be addressed through separate documents, such as compilation guidance and clarification notes, in addition to the current BPM6 and 2008 SNA. In other cases, the issues may go to the heart of the national and international accounts, and can only be addressed within an updated version, "benchmark" type of revision, of the international standards. Furthermore, issues that are viewed as affecting fundamental principles may be best addressed through "extensions" or supplements to the basic, core concepts (e.g., extending the production boundary to include unpaid household service work such as cleaning), providing alternative concepts, but leaving the "core" framework intact.

At the same time, the solutions to be provided should be tested by several critical restrictions. Among the most important are: (i) statistical compilation feasibility; (ii) data source availability and accessibility; (iii) objectivity; (iv) and flexibility to implement second-best solutions, due to differences across countries related to their economic structure and statistical capacity. As such, the conceptual guidance should provide a roadmap for what can be done by national statistical offices and central banks, considering existing as well as new and innovative good statistical compilation practices.
The following questions are worth considering:

- What areas, economic and financial developments are not sufficiently or well covered?
- What issues would change the fundamentals of the system? Do benefits of the potential change outweigh costs?
- What issues can be addressed through granular, supplementary, or extended measures?
- Are issues relevant for most countries? Could the potential solutions be implemented by national statistical compilers?

The major priority issues discussed below are broadly consistent with issues put forward by other international organisations. The only addition in which we seek new efforts in the next generation of standards is the *Informal Economy*, where the concept is being tested with new types of jobs largely linked to digitalization, e.g. the gig economy. Furthermore, practical measurement issues on the informal economy need to be addressed as well. In fact, most of the priorities are substantially linked: globalization, digitalization, informality, income and wealth inequality or the need for welfare indicators are often closely linked and reinforce each other.

3. Globalization

The broad topic. Economic globalization has led to tighter integration of economies, with the creation of global value chains, allowing businesses to organize their production chains more efficiently. Furthermore, multinational enterprises (MNEs) organize their production and marketing at a global level and may utilize special purpose entities (SPEs) structures, to obtain benefits from different legal and tax regimes which allow them to maximize profits after taxation. In addition, as the economy becomes increasingly digitalized, digital businesses can be "born global" reaching customers all around the world at an unprecedented scope and scale. This has increased the complexity of compiling economic statistics as it is more difficult to break down production activities on a country-by-country basis as well as to determine in which cases there is a real transfer of economic ownership as opposed to the provision of (manufacturing or other) services. As a result, the measurement of key economic indicators, including GDP, has been affected.

These emerging and ever-changing global production arrangements pose challenges to macro-economic statistics. The IMF's Committee on Balance of Payments Statistics (BOPCOM), have made progress, described below, on certain issues that can be addressed through more supplemental and granular statistics (that do not change the core fundamentals of the accounts) to shed light on the phenomena.

Globalization: Progress made so far. SPEs have evolved beyond those structures anticipated in the current statistical manuals. While originally SPEs were mostly set up by financial institutions, they have evolved to include nonfinancial specialized entities established by MNEs to manage intellectual property rights, research and development, trade, and other activities as part of the group-wide financial and profit maximization strategy. The common denominator of these activities is often tax arbitrage among jurisdictions in a context of free capital movements. Considering the evolving nature of SPEs, the IMF BOPCOM's Task Force on SPEs proposed an international definition of SPEs in the context of cross-border statistics, providing further guidance on SPEs beyond what is in the current statistical manuals which allows for cross-country comparable data to be collected.[2] This definition was adopted by BOPCOM and supplementary SPEs data collection will begin by the end of 2021.

[2] https://www.imf.org/external/pubs/ft/bop/2018/pdf/18-03.pdf

Additional statistics are also needed to better analyze global value chains (GVCs), including the role of MNEs in these processes. Key to better understanding this fragmentation of production is interpreting current trade related statistics as gross exports when they often contain significant foreign intermediate goods and services. Since GVCs are not evident in the gross trade data, potentially distorting who really trades with whom, current measures of bilateral trade balances may need supplemental measures for optimal policy analysis requiring national statistics to offer additional data to shed light on the partial globalization of their respective national economies. It is also necessary to better identify the role of MNEs in current account transactions, potentially identifying intra-MNE flows some of which utilize transfer pricing to shift profits to minimize an MNEs global tax burden. Breaking down the goods and services account by enterprise characteristics, such as industry, nationality (foreign vs. domestic) and firm size would be highly useful for analysis of globalization issues. However, is it feasible, and if so to which degree of granularity?

Globalization: Issues for further consideration. While progress has been made, further research is needed in certain areas. One such issue is addressing the globalization issue through extensions of the current BPM6 framework. To do this, the BPM6 framework could be extended to provide an alternative view that complements the residency concept, i.e. territorial, on which the fundamentals of BPM and SNA are based. Introducing a nationality concept, i.e. ownership, would allow users to better understand the passage from GDP to gross national income or the value of conventional international investment position statistics. Further exploration on how to manage the duality of residence/nationality within the international accounts' framework would be needed before such extensions could be produced.

However, some issues– in which further research is still needed– may need to be addressed through changes in the core framework as advances in technology and communication, dominance of MNEs, use of SPEs, increasing free capital movements, as well as the distortion of GVCs themselves due to profit-shifting are seriously testing some of the fundamentals of the system.

It can be argued whether SPEs should be considered as separate institutional units from their parent. An intrinsic difficulty to record and treat appropriately SPEs is when their creation is due to tax-arbitrage and profit-shifting, as is often the case. Currently, the residence of an SPE is determined according to the economic territory under whose legal jurisdiction the entity is incorporated or registered. If the entity is legally located in an economy different from its parent, then it is recognized as a separate institutional unit. As discussed in Moulton and Van de Ven[3], there are two reasons why: (i) this

[3] http://papers.nber.org/conf_papers/f100570.pdf

treatment is consistent with the actual cross-border cash flows resulting from economic transactions; (ii) rerouting such transactions would require the exchange of individual data between statistical offices, which is not possible given existing legal constraints.

Another challenge is determining the economic ownership of intellectual property products (IPPs), e.g. research and development, because of their intangible nature their ownership, location and use are not easily observed. Key amongst this is trying to determine economic ownership of these intangible assets amongst an MNE group. The parent may assign legal ownership of IPPs to SPEs which otherwise do not contribute to the MNE´s production activities. Recent work of the IMF's Task Force on SPEs put forward a proposal to collect supplementary information on SPEs for a reduced number of BOP and IIP components beyond direct investment activities. As transactions in goods would be relevant for merchanting SPEs, a separate line for net merchanting by SPEs is required. Regarding services, four distinct components of services have been included in the reporting list where SPEs can be of relevance: transport, financial services, charges for the use of intellectual property, and other business services.

The principles of economic ownership of IPPs are difficult to apply inside MNEs because the use of the IPP by one part of an enterprise group does not prevent the simultaneous use by another part and legal ownership of IPPs can be placed anywhere amongst the group. Currently, the guidance states to record a change in economic ownership when a financial transaction between two institutional units occurs, which corresponds to a change in legal ownership. Various options for deviating from this suggested approach[4] have been discussed but further research is needed as to what can be done in practice. Furthermore, the recording of what entity is the economic owner of the IPP also has implications for how and where the related charges for the use of the intellectual property are recorded. Since IPPs are often considered corporate property, payments for its use may not always be observed separately and may instead be shown as distributed or retained earnings from foreign direct investment.

In the context of increasing globalization, external sector surveillance has become central to the Fund's mandate. Monitoring and assessing global imbalances (large current account and net international investment position imbalances) and disentangling any measurement distortions are of the highest importance. IMF research suggests that statistical measures of current accounts may not always give an accurate picture of true external imbalances and one such measurement distortion is the treatment of retained earnings on portfolio equity. Currently, retained earnings on portfolio equity are not

[4] http://www.iariw.org/copenhagen/konijn.pdf

distributed to foreign portfolio investors and therefore, not distributed to the owner. This is inconsistent with the treatment of retained earnings of a foreign direct investment enterprise that are treated as if they were distributed to foreign direct investors in proportion to their ownership of the equity of the enterprise. It has been proposed that this treatment could be extended to portfolio equity investment to help alleviate the measurement distortion. However, a careful consideration of its implications and feasibility is warranted before committing.

4. Digitalization

The broad topic. Digitalization encompasses a wide range of new applications of information technology in business models and products that are transforming the economy as well as welfare implications. The effects of this technological innovation include sharply lower prices and increased efficiency in computers, smart phones, and the Internet; new goods and services; innovations in financial markets and new methods of payment (e.g. Fintech); and reductions in costs and improvements in quality and efficiency associated with the use of technology.

Digitalization: Progress made so far.

There is no generally agreed definition of the "digital economy" and consequently a common metric to measure its size. While digitalization has penetrated many activities, and almost the entire economy is affected-mostly positively by digitalization, it is more realistic to focus the measurement of the digital economy on a concrete range of economic activities at the core of digitalization. The IMF (2018a) Executive Board paper on *Measuring the Digital Economy*[5] distinguishes a "digital sector" that covers the core activities of digitalization, ICT goods and services, online platforms, and platform enabled activities. Exercises such as these better highlight the digital economy without making changes to the fundamental principles. Furthermore, digitalization also raises measurement issues and new data needs for external, monetary and financial sector statistics. Improved measurement of digital products and transactions could improve measurement of inflation and balance of payments as many of the online platforms are not resident in the economy of the user. Recent work by the OECD, with IMF participation, on Measuring Digital Trade and Digital Supply-Use Tables will provide more granular data to better shed light on digital transactions and products based on an agreed framework.

Another area in which progress has been made is clarifying the statistical recording of crypto assets in the international standards. Crypto assets

[5] https://www.imf.org/en/Publications/Policy-Papers/Issues/2018/04/03/022818-measuring-the-digitaleconomy

combine properties of currencies, commodities and intangible assets. The IMF (2019a) recently released a clarification note on *The Treatment of Crypto Assets in Macroeconomic Statistics*[6] that presents an overview of crypto assets and provides guidance on the classification for their treatment in macroeconomic statistics based on the current statistical standards and classifications. At present, Bitcoin-like crypto assets do not meet the definition of a financial asset, so they should be classified as produced nonfinancial assets as a distinct sub-category under valuables.

***Digitalization: Issues for future consideration*. *The value of (digital) data*.**

While much can be addressed by providing more granular or supplementary statistics, there may be a need to go beyond the current international standards to better account for the role of data in the modern digital economy. Data have always had a central role in business decision making. Businesses strive to gather data on customers, to improve products and processes to enhance productivity, improve performance, and increase profitability. As storage and acquisition costs decrease and processing capacity (software, IT hardware) increases, an explosion in data accumulation is happening. The simple fact that data is in electronic form allows it to be analyzed for insights and decision-making at an unprecedented scope and scale. In some sense, data itself has been transformed: it has become digital data. This digital data has allowed for new information/knowledge creation that could not have been done if the data were not in digital form. While initial work by the IMF (2019b)[7] has begun, there is not yet any agreement on the statistical treatment of data, thus further research is needed to determine if data is an asset and, if so, is it produced or non-produced; as well as how to determine its value.

5. Economic Welfare

The 6th IMF Statistical Forum (Measuring Economic Welfare: What and How?) held in November 2018 helped set the framework for what measures "Beyond GDP" may be needed to better understand changes in economic welfare in the digital age as GDP is not intended to be a comprehensive measure of economic welfare. In particular, how the gap between welfare growth and growth of real private consumption, a measure of economic welfare, is widening in the presence of new and rapidly changing, often free, digital services. Furthermore, as governments and international organizations, such as the IMF, adopt inclusive growth strategies, there is a clear need and expectation to go beyond measuring the size and aggregated growth of the economy to understand how the benefits of economic activity are being

[6] https://www.imf.org/external/pubs/ft/bop/2019/pdf/Clarification0422.pdf

[7] https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.20/2019/mtg1/IMF.pdf

distributed. Further work on linking households' indicators from sectoral accounts, micro data and macro aggregates should be high for national and international statistical agencies.

In addition, while the SNA provides a brief general discussion of real income reflecting the real purchasing power, it is not prescriptive of what types of income measures would be needed to appropriately measure economic welfare. In this respect, a discussion on real household (adjusted) disposable income as being a more appropriate measure of household's economic well-being instead of real GDP could be emphasized to gain traction in countries that do not publish income measures. Furthermore, the SNA could be more prescriptive on the choice of deflator, emphasizing that the appropriate measure of real income would use a deflator containing prices for the uses of that income rather than from the sources of the income.

6. Informal economy

The informal economy broadly defined should comprise (i) the production of goods and market services of households; and (ii) the activities of corporations (illegal; underground) that may not be covered in the regular data collection framework for compiling macroeconomic statistics. Furthermore, the gig economy, facilitated by digitalization, has given a new meaning to the concept of informality in the economy, but also may provide scope for better measuring these activities using big data. The current statistical manuals, 2008 SNA and BPM6, fall short in providing a clear definition that is suitable for compiling the requisite statistics for policy analysis; therefore, more clarification in the international standards may be needed. To make progress on this issue, the 7th IMF Statistical forum in November 2019, will focus on estimation methods by identifying good, feasible, traditional compilation methods as well as innovative data sources from big data to the use of drones–used to measure agriculture and construction activity– to better measure the informal economy.

7. Discussion and Conclusion

A considerable number of current statistical challenges may be satisfied by providing more granular or supplementary information, allowing the flexibility needed. Overall, the basic structure of the current international statistical standards remains valid and there is no need to propose major changes to the core macroeconomic statistics framework. Furthermore, given that many countries have not yet adopted the 2008 SNA[8] and are unable to comply with the minimum required scope and detail of national accounts data, it is,

[8] Only 88 countries had adopted the 2008 SNA by December 2018. https://unstats.un.org/unsd/statcom/50thsession/documents/2019-8-NationalAccounts-E.pdf

therefore, not an urgent need for many countries to update the international standards. Nevertheless, developments have continued to progress especially in the areas of globalization, digitalization, economic welfare, and the informal economy that need to be systematically tackled, when thinking about the next generation of international statistical standards, a discussion to which this paper tries to contribute.

References

1. IMF, Balance of Payments and International Investment Position Manual, Sixth Edition, 2009
2. IMF (2018a), Measuring the digital economy, IMF Executive Board paper
3. IMF (2018b) BOPCOM's Task Force Special Purpose Entities Final Report
4. IMF (2019a), The treatment of crypto assets in macroeconomic statistics, IMF Clarification Note 5. IMF (2019b), Measuring the digital economy in macroeconomic statistics: the role of data, Meeting of the Group of Experts in National Accounts April 2019
5. Moulton, B., Van de Ven, P., Addressing the challenges of globalization in national accounts, National Bureau of Economic Research Conference March 2018
6. Ribarsky J., Konijn P., Nijmeijr H., and Zwijnenburg J. (2018) "Measuring the Stocks and Flows of Intellectual Property Products". International Association for Research in Income and Wealth Conference August 2018
7. United Nations, European Commission, IMF, Organisation for Economic Co-operation and Development, World Bank, System of National Accounts (SNA) 2008, New York, 2009

It's about time for a change: Broadening the SNA framework to account for sustainability and well-being

Peter van de Ven
Organisation for Economic Co-operation and Development (OECD)

Abstract

Economic growth, or the volume growth of GDP, can be considered as the single most successful macro-economic indicator, having showed its policy relevance since the depression of the 1930s. For many decades, it has basically defined whether the economy is successful or not, or even more broadly whether societal developments are going in the right or wrong direction. As such, it has largely driven, and continues to drive, the policy agenda. GDP actually does a pretty good job as an indicator of (monetary) economic activity. However, when it comes to monitoring well-being of people, including the future sustainability of this well-being, economic growth has many fallacies and caveats. The link between continuous increases of GDP and enhancing (sustainable) well-being is more and more questioned, debated and considered totally flawed. From an environmental perspective, navigating on GDP alone may show to be the shortest route to disaster.

For the above reasons, there is an increasing user demand for arriving at better metrics that provide a more encompassing measure of developments in (sustainable) well-being. It may not be possible to find what is considered by some as the holy grail, an alternative catch-all indicator, that provides a perfect monitoring instrument for well-being, which also takes into account the present-day losses (or gains) in the possibilities to generate future well-being. The pursuit of such an indicator may show to be a dead end road. Well-being is a multi-faceted phenomenon that may only be captured by a dashboard of indicators, such as for example the OECD Better Life Index.

The paper argues in favour of trying to define and populate an underlying conceptual and statistical accounting framework for the indicators which are typically considered relevant for well-being and its future sustainability. A further enrichment of the central framework of national (monetary) accounts, by including and combining a standard range of accounts on environment, health, education and time use, could provide an excellent tool for enhanced policy analysis. All of these additions need not necessarily be defined in monetary terms, although monetising (degradation of) ecosystems could further support the urgency of dealing with environmental sustainability.

Keywords

Economic growth; national accounts; OECD Better Life Index; sustainability; well-being

1. Introduction

Kenneth Boulding once noted that "anyone who believes exponential growth can go on forever in a finite world is either a madman or an economist"[1]. Amongst others, Philipsen (2015) shows the ridiculousness of continuously pushing for a 3% growth of world GDP, which would result in a doubling of the world economy every quarter of a century, and leading to a world economy which by the end of the 21st century would be eight times larger than the current one. Adding another century would lead to a 128-fold multiplication of the current level of economic activity. All of this is not to say that compiling GDP-numbers is pretty much useless. Clearly, monitoring and analysing economic activities are important in their own right, for example to support policies for designing a financially sustainable economy. But that should not lead to policies that continuously and exclusively beat the drum of an unconditionally higher GDP. For what purpose? For whom? Economic growth cannot be the ultimate objective of a society. As many have said, we need a better navigation system that guides policy towards the enhancement of well-being of people, without jeopardising the sustainability of well-being for future generations to come. But often voices become much softer, or even silent, when it comes to concrete alternatives which could provide clearer guidance for the future direction of societal developments, have a rigorous and conceptually sound underlying measurement framework, and – last but not certainly least – are easy to communicate.

Whatever the case, the quite alarming societal developments call for further action to arrive at better metrics that provide a more encompassing measure of developments in (sustainable) well-being. It may not be possible to find what is considered by some as the holy grail, an alternative catch-all indicator, that provides a perfect monitoring instrument for well-being, which also takes into account the present-day losses (or gains) in the possibilities to generate future well-being. The pursuit of such an indicator may show to be a dead end road. Well-being is a multi-faceted phenomenon that may only be captured by a dashboard of indicators, such as for example the OECD Better Life Index; see http://www.oecdbetterlifeindex.org/.

This short paper includes a proposal to broaden the traditional framework of national accounts into an accounting framework that can support the analysis of dashboards of indicators for measuring well-being and sustainability. It thus tries to provide a linking pin between the current set of national accounts and the dashboards of indicators. After a short description, in Section 2, of initiatives closely related to the current system of national

[1] United States Congress, House (1973) Energy reorganization act of 1973: Hearings, Ninety-third Congress, first session, on H.R. 11510, page 248.

accounts, Section 3 of this paper goes one major step further, by presenting an outline of such a broader accounting framework, which is considered practically feasible in a relatively short time period. Section 4 concludes with some suggestions on the way forward in implementing such an approach.

2. Current initiatives closely related to the current system of national accounts

One of the most influential initiatives to arrive at a better understanding of well-being is the "Report by the Commission on the Measurement of Economic Performance and Social Progress" by Joseph Stiglitz, Amartya Sen and Jean-Paul Fitoussi (Stiglitz et al, 2009). The report contains various recommendations, among which the first five are directly related to macro-economic statistics, as follows:

1. When evaluating material well-being, look at income and consumption rather than production.
2. Emphasise the household perspective.
3. Consider income and consumption jointly with wealth.
4. Give more prominence to the distribution of income, consumption and wealth.
5. Broaden income measures to non-market activities.

The main thrust of these five recommendations is to not only look at (developments in) GDP, but also to household disposable income, the distribution of income, consumption and wealth, and the free services provided by unpaid household activities (taking care of children and elderly people, cooking meals at home, cleaning, etc.). As such, the Report also recognises that the system of national accounts is much more than GDP alone, a mistake that is often being made, if not explicitly then at least implicitly. The system of national accounts contains a complete and consistent set of accounts that describe all (economic) transactions and positions for households, corporations, government, etc.

Importantly, the Report does not contain a recommendation to objectively capture well-being in a single metric, by for example monetising all aspects that have a positive or negative impact on well¬being. The same holds for trying to capture (environmental) sustainability, by monetising all negative externalities from economic activities on the environment. The Report considers well-being as a multi¬dimensional phenomenon, and preference is given, at least for the time being, to define the various aspects that affect well-being, and then select indicators for monitoring the developments for each of these aspects. As a follow-up to the 2009 Report, a High Level Expert Group (HLEG), the secretariat of which was provided by the OECD, was set up in 2013 to continue the work of the Stiglitz-Sen-Fitoussi Commission. Their report "Beyond GDP. Measuring what counts for economic and social performance" (see Stiglitz et al., 2018) makes, amongst others, a plea for the "capital approach", in which progress is monitored by the development of economic, human and environmental capital, and by the vulnerability and resilience of systems.

In response to the Stiglitz-Sen-Fitoussi Report, and also motivated by the OECD Inclusive Growth Agenda, among which the work on broader measures of well-being (see e.g. the OECD Better Life Index), several initiatives have been taken forward in the area of national accounts as well, with the objective to put more focus on (the distribution of) household disposable income, instead of GDP:

- The dissemination of a dashboard on households' economic well-being; see http://www.oecd.org/sdd/na/household-dashboard.htm.
- The dissemination of a quarterly news release on "growth and economic well-being"; see e.g. http://www.oecd.org/sdd/na/Growth-and-economic-well-being-oecd-02-2019.pdf.
- A working paper on the decomposition of differences between GDP growth and growth in real disposable income.
- A working paper on the impact of valuing unpaid household activities.
- Aligning micro data on the distribution of income and consumption to national accounts, to arrive at distributional information that is consistent with macro-economic indicators; see e.g. the latest working paper on this topic.

Furthermore, much effort has been put into the further development and dissemination of environmental-economic accounts. These accounts are based on the System of Environmental-Economic Accounting – Central Framework (SEEA-CF), which the UN Statistical Commission endorsed in 2012 as a global standard for compiling such accounts. In addition to an extended accounting for natural assets, and perhaps more importantly from a monitoring perspective, the SEEA-CF includes a set of physical flow accounts, in which natural inputs, products and residuals are linked to economic activities. Physical supply and use tables are included for energy, water, and various material flows (emissions to air, emission to water, and solid waste). Another part of the framework concerns a more prominent accounting for environmental activities, by identifying economic transactions within the system of national accounts which mainly relate to *"… economic activities whose primary purpose is to reduce or eliminate pressures on the environment or to make more efficient use of natural resources*" (§ 1.30 of SEEA-CF). A final set of accounts breaks out environmental taxes, subsidies and similar transfers.

The uptake of implementing SEEA-consistent accounts has been very good, with currently 69 countries having programmes on environmental-economic accounting. The goal for 2020 is to have at least 100 countries with ongoing, well-resourced programmes in line with SEEA-CF. All of this is a major step forward, which will allow for a better monitor and analysis of the externalities of economic activities in the form of emissions, and of the growth of environment-friendly activities. Furthermore, an improved accounting for mineral and energy resources would make it possible to calculate a Gross Domestic Product, which is not only adjusted for depreciation of produced assets, but also adjusted for depletion of natural resources.

However, much more needs to be done. The economy and the society at large are embedded in and depending on the limitations provided by Planet Earth. From a sustainability perspective, the most important assets are ecosystem assets. § 2.31 of SEEA 2012 – Experimental Ecosystem Accounting (SEEA-EEA) defines ecosystems as *"... spatial areas comprising a combination of biotic and abiotic components and other characteristics that function together"*. These assets provide ecosystem services, benefits used in economic and other human activity, a rather euphemistic formulation for services on which human and other life depends. In SEEA-EEA, three main types of services are distinguished: (i) provisioning services (e.g. timber from forests); (ii) regulating services (e.g. forests proving carbon sinks); and (iii) cultural services (e.g. the pleasure of visiting a national park). However, ecosystems accounting is not straightforward at all. In physical terms, consensus has more or less been achieved on the way forward. However, accounting for the monetary value of the stocks of ecosystem assets, and their degradation over time, is a slightly different story. Notwithstanding the complexity, much progress has been made in recent years, and work is ongoing to further improve methodologies. In future, all this work could potentially lead to the compilation of physical and monetary estimates for stocks and degradation of ecosystems.

3. A broader framework of "national accounts"

Although the above initiatives are important in their own right, it is of the utmost importance to develop metrics that cast a wider net on the monitoring of the well-being of people and the sustainability of societal developments. As (sustainable) well-being is a multidimensional phenomenon, it may not be possible to capture it in one catch-all indicator, and one thus has to agree and rely on a set of indicators which monitor the most relevant aspects. An important example of capturing well-being by a limited set of indicators is the OECD Better Life Index, in which eleven areas of (sustainable) well-being are being monitored: housing; income; jobs; community; education; environment; civic engagement; health; life satisfaction; safety; and work-life balance. For each of these aspects, regional and interpersonal distributions are also taken into account. If inequalities of say income, wealth, health, education, etc. coincide, the impact on well-being for the relevant people at the bottom of these distributions can be very detrimental indeed.

It is proposed here to take all of this one step further, by developing a broader accounting framework that supports the monitoring and analysis of the interrelations between the various aspects of well-being, thus providing a better understanding of the trade-offs and the win-wins between the various domains. For example, what's the relationship between on the one hand the output of the medical industry and unpaid household activities on care for (non-)household members, and on the other hand the health outcomes of people, and how does this affect, for example employment and government finance. How to improve health outcomes? Should we spend more money on prevention, on development of pharmaceuticals, on improving medical techniques, and how much money are we willing to spend? To answer this type of questions and to provide a link with relevant outcome indicators, one may consider supplementing the traditional set of national accounts data on the production of medical goods and services and government finance, with more detailed business statistics on the medical industry, granular administrative data on medical treatments, and hours spent on relevant time use categories. In the first stages of development, it may not be possible to define a framework that covers all indicators, and for reasons of practical feasibility, it is advisable to limit the framework to a subset of indicators; see also below.

This way of thinking about linking various areas of statistics has been developed and implemented in the area of environmental sustainability (see e.g. the above on SEEA), but there are other promising initiatives as well. For example, quite a number of countries have done work on health satellite accounts, while more and more work is being done on exploiting administrative data on medical treatments, to arrive, for example, at better measures for the volume changes in the provision of health services. Furthermore, in the area of education and training, a conceptual framework for compiling satellite accounts[2] for education, training and human capital has been developed; see UNECE (2016). The same is true for satellite accounts on unpaid household activities (see UNECE, 2017), where various attempts have been made to integrate numbers from the time use surveys with the hours worked in paid employment, and to value the services provided within households. In respect of the latter, a framework, which combines the traditional national accounts with more granular and more timely data on time use, would provide an excellent tool for monitoring and analysing shifts in time spent between paid activities, unpaid household activities and leisure time. Such shifts are considered highly relevant for the measurement of people's well-being, whether this may concern activities benefiting people's own well-being (e.g., work-life balance, leisure), other people's well-being (e.g., childcare, eldercare, volunteering), or both (e.g., socialising). Furthermore, it would be great to have additional information on e.g. digital activities, such as time spent on social media, search activities, etc., thus allowing for an alternative way of measuring consumer surplus that may arise from the digitalisation of the society[3].

More generally, as a point on the horizon, one would like to see the development of an overarching accounting framework, in which statistics on economic, societal and environmental issues are integrated (not necessarily monetised), and in which one can easily drill down into micro-datasets. It is clear that this can only be a long-term goal, also requiring the development of a suitable conceptual framework. As a more realistic goal for the nearer future, one could envision the regular compilation of certain thematic satellite accounts, such as the ones mentioned in the above. Having accounts for the environment, health, education and unpaid household activities, or time use more generally, compiled on a regular basis for a substantial number of countries would definitely support the monitoring and analysis of quite a number of well-being aspects included in the OECD Better Life Index, and other dashboards at international and national level. In doing so, it is not necessary to compile all accounts at a quarterly or annual basis. Some accounts, for which structural developments are the primary focus, one could think of a compilation every 2-3 years, depending on user demands and the availability of source data.

In developing such a broader framework, one should acknowledge the importance of communication. Referring to the traditional set of national accounts as being the "central framework" or the "core" set of national accounts, and referring to the measurement frameworks for other areas as being satellite accounts, is not particularly helpful. In line with Vanoli (2017), we need to rethink terminology and the content of what's currently being referred to as the central framework. Vanoli proposes to refer to the current set of national accounts as the System of National Economic Accounts (SNEA), and to include a much broader set of accounts in the central framework of national accounts. One would perhaps even want to go a step further when it comes to the term for the current set of national accounts,and refer to it as the System of National Monetary Accounts, to make clear that economy is more than a consistent set of monetary transactions and positions.

---

[2] A satellite account is a set of accounts that are linked to, but distinct from, the central framework of national accounts. One type involves some rearrangement of central classifications and the possible introduction of complementary and/or more detailed elements, which do not change the underlying concepts of the SNA in a fundamental way. A second type of satellite analysis is mainly based on concepts that are alternatives to those of the SNA. These include a different production boundary, an enlarged concept of consumption or capital formation, an extension of the scope of assets, and so on.

[3] It won't be possible to achieve the requested granularity and timeliness by applying traditional survey methods; more thinking is needed about the potential of big data to compile time use data, or at least to supplement current data collection methods.

Furthermore, Vanoli also presents a concise conceptual foundation for the broader set of accounts, with reference to four spheres and their related information systems: economy, people, nature and society. A similar plea for a new and comprehensive "System of Global and National Accounts", including an in-depth analysis and description of how such a system should look like can be found in Hoekstra (forthcoming). Hoekstra argues for a distinction of four sets of interrelated accounts, three describing the environment (Global Environmental Accounts), the society (Global Societal Accounts) and the economy (Global Economic Accounts), and one describing distributional aspects (Global Distribution Accounts). A separate set of Global Quality Accounts is distinguished for the definition and recording of key indicators that can be derived from the other three accounts, to provide a summary on whether things are moving in the right direction.

4. The way forward

The development of a full-scale conceptual framework for capturing all aspects of (sustainable) well-being may be too demanding and time-consuming. Here, a more pragmatic approach is being proposed, to get things moving forward relatively quickly. Relatively quickly, as following this approach, it may still take quite some time to define the conceptual framework, including the templates, discuss these proposals and have them endorsed at the international level, and – last but certainly not least – to get all of this implemented by countries.

In its 49th meeting held on 6-9 March 2018, the UN Statistical Commission agreed to develop guidance notes on three priority areas for which further clarifications and guidance are needed in the context of the further development of the 2008 System of National Accounts: (i) digitalisation, (ii) globalisation, and (iii) well-being and sustainability. For each of these areas, dedicated Task Forces have been set up in the first half of 2019, with the goal of delivering draft guidance notes in the course 2020. The Terms of Reference of the Task Force working on well-being and sustainability includes, in addition to unpaid household activities, environmental-economic accounting, and distribution of income, consumption and wealth, the following explicit reference: "defining a broader framework for capturing economic activities, well-being and sustainability". In moving forward quickly, one would prefer to keep the group of people actively contributing to the drafting of the guidance notes relatively small, but it is also considered of the utmost importance to involve, at some stage, specialists from other areas of expertise (environment, social issues, education, health, time use, etc.) as well. One should avoid at all costs that the development of this broader framework of "national accounts" is looked upon as a form of economic imperialism. The objective is to arrive at a consistent framework, which covers much more than the economy alone.

References

1. Hoekstra, R. (forthcoming), "Replacing GDP by 2030. Towards a common language for the wellbeing and sustainability community", Cambridge University Press, Cambridge, forthcoming.
2. Philipsen, D. (2015), "The little big number. How GDP came to rule the world and what to do about it", Princeton University Press, Princeton, 2015.
3. Stiglitz, J.E., J-P. Fitoussi, and M. Durand (2018), "Beyond GDP. Measuring what counts for economic and social performance", Paris, 2018. Available at: http://www.oecd.org/publications/beyond-gdp-9789264307292-en.htm
4. Stiglitz, J. E., A. Sen, and J-P. Fitoussi (2009), "Report by the Commission on the Measurement of Economic Performance and Social Progress", Paris, 2009. Available at: http://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report.
5. UN, European Commission, IMF, OECD, World Bank (2009), "System of National Accounts 2008", New York, 2009. Available at: https://unstats.un.org/unsd/nationalaccount/docs/sna2008.pdf.
6. UN, European Commission, FAO, IMF, OECD, World Bank (2014a), "System of Environmental-Economic Accounting 2012 – Central Framework", New York, 2014. Available at: https://seea.un.org/sites/seea.un.org/files/seea_cf_final_en.pdf.
7. UN, European Commission, FAO, IMF, OECD, World Bank (2014b), "System of Environmental-Economic Accounting 2012 – Experimental Ecosystem Accounting", New York, 2014. Available at: https://seea.un.org/sites/seea.un.org/files/seea_eea_final_en_1.pdf.
8. UNECE (2016), "Guide on Measuring Human Capital", New York and Geneva, December 2016. Available at: https://www.unece.org/fileadmin/DAM/stats/publications/2016/ECECESSTAT20166_E.pdf.
9. UNECE (2017), "Guide on Valuing Unpaid Household Service Work", New York and Geneva, December 2017. Available at: https://www.unece.org/fileadmin/DAM/stats/publications/2018/ECECESSTAT20173.pdf.
10. Vanoli, A. (2017), "The Future of the SNA in a Broad Information System Perspective", The Review of Income and Wealth, series 63, supplement 2, December 2017, page S238 – S265. Available at: https://onlinelibrary.wiley.com/doi/epdf/10.1111/roiw.12332.

Broadening the scope of GDP – a stocktaking on including ecological developments

Albert Braakmann
Federal Statistical Office of Germany, Wiesbaden, Germany[1]

Abstract
Gross Domestic Product (GDP) is the most prominent indicator of the System of National Accounts (SNA), which is widely used for policy and administrative purposes. GDP in real terms (i.e. price adjusted) has been accepted since some decades as a measure for economic growth. At the same time GDP often is considered as an indicator for societal progress and well-being. However, as highlighted by the famous Stiglitz-Sen-Fitoussi-report of 2009, such an interpretation of GDP is biased and can be misleading, given that most social and environmental cost are not covered, or even worse, their repair adds to GDP. In the case of environment the key words in this context are ecosystem services, depletion and degradation and as a practical example the repercussions of the global climate change could be mentioned. This paper therefore looks at the conceptual as well as empirical pros and cons of broadening the concept of GDP by including ecological developments.

Keywords
National accounts; Environment; Ecosystems; Well-Being.

1. Economic and environmental accounts

Air and water are indispensable requirements to human life, the cleaner the better. The availability of food largely depends on environmental conditions. Furthermore, natural resources are important for our economies: The production and consumption of goods and services depends on extracting minerals and other resources from nature. Eventually the material circle ends by returning waste to nature.

For many economic decisions the yardstick mostly used is economic growth measured by real gross domestic product (GDP), i.e. volume of goods and services available to human societies. This key aggregate of the worldwide System of National Accounts (SNA)was conceived as a measure for the production of goods and services and often is taken as an indicator of societal progress. In this context the production boundary is crucial, since it defines the scope of production activities to be covered by GDP. Excluded for instance is housework for the own household, although in most cases these activities

[1] The views expressed in this paper are those of the author. They do not necessarily reflect the views of the German Federal Statistical Office.

add substantially to well-being. Most social and environmental costs coming along with economic growth are not deducted, or even worse, the repair activities increase GDP. Excluded furthermore are natural processes not under the control of resident economic units.

Looking at the GDP growth rates seems to suggest that unlimited economic progress is possible. Opposed to this it has been realized since some time, that the world and its resources are limited. This leads to the question, whether we can continue like today or whether there is a need to consider a more sustainable development. The issue of sustainability is further increasing by the worldwide population growth: According to United Nations forecast there will be over 10 billion people living on earth in 2055, representing a population increase of more than 30 % in just 25 years. The challenges ahead imply to think about developing adequate (additional) statistical indicators.

Sustainability is the core element in the resolution "Transforming our world: the 2030 Agenda for sustainable development" adopted by the General Assembly of the United Nations (UN, 2016). But already several decades before the publication of the report of the Club of Rome on the limits to growth (Meadows, D., 1972) initiated discussions at various levels of societies all over the world. From these discussions the idea of a more comprehensive approach emerged. Different strands and indicators were proposed. Developing a so-called green GDP, i.e. an environmentally adjusted GDP, was an immediate approach. In addition, the 1987 report of the Brundtland Commission "Our Common Future" (World Commission on Environment and Development, 1987) focused on the links between economic, societal and environmental development.

At the statistical level, research activities increased fostering integrated environmental and economic accounts. Briefly, this led to the publication of a corresponding UN-handbook in 1993, updated in 2003, ultimately revised and published as System of Environmental-Economic Accounting 2012 - Central framework, which in March 2012 was adopted as an international statistical standard (SEEACF, 2012).

In general terms, the SEEA-2012-CF highlights the relationship between economy and environment by looking at "natural inputs" from the environment to the economy, e.g. mineral resources, timber, water or energy. On the other hand "residuals" flowing from the economy to the environment, e.g. solid waste, air emissions, return flows of water are considered as well (SEEA-CF, 2012, p. 13). These flows usually are compiled in physical terms using a supply-use framework known from national accounts. In addition to such flows, the methodology for compiling environmental assets are covered, where the focus is on individual components like mineral and energy resources, timber, water resources and land, both, in physical and value terms.

As a third item, certain environmental activities, like protection expenditures, are included in the SEEA-2012-CF.

However, accounting for degradation and other measurement topics associated with ecosystems are not covered in the SEEA-central framework (SEEA-CF, 2012, p. ix). Excluded as well are oceans and the atmosphere, due to their magnitude. Full valuation of assets and flows related to natural resources and land beyond the valuation included in the SNA remains an outstanding issue (SEEA-CF, 2012, p. viii). Basically the SEEA-2012-CF focusses on the material benefits from the direct use of environmental assets or natural inputs for the economy by enterprises and households. By contrast, this does not cover the non-material benefits from the indirect use of environmental assets, like benefits from ecosystems services (SEEA-CF, 2012, p.13).

2. Ecosystem services and assets

2.1 Conceptual approach

A more comprehensive and growing approach to look at the role of nature for sustainable economic growth and human well-being is based on the concept of ecosystem services and assets[2]. In general terms the idea is that ecosystems (ES) provide benefits to humanity free of charge, which have a fundamental importance for human well-being, health, livelihood and survival. (MA, 2005; TEEB, 2010). Often-quoted ES-examples are cleaning of air and water, wild food and crops, natural medicines, but also pollination, climate regulation, filtration of pollutants by wetlands, recreation facilities, soil formation and photosynthesis or biodiversity (MA 2005, TEEB 2010, p. 7). The amount of benefits provided to human societies usually depends on the type and quality of the ecosystem considered and hence may vary from one region to another. The spatial perspective therefore is a fundamental element of the ES-concept.

The ES-approach goes a step further and looks at the natural processes from an economic viewpoint. Like in the SNA, a distinction is made between flows, which are measured over a period of time and stocks, which are captured at a certain point in time:

- The benefits provided by ecosystems to human societies are interpreted as services, in this case ecosystem services. The natural processes from which the ES services are resulting are looked at as a kind of production activity.[3]
- On the other hand, the ecosystem capacity is considered as an asset, sometimes called natural capital. The ES capacity may be negatively or

[2] The experimental methodology for ecosystem accounting has been published separately (SEEA-CF, 2012).

[3] However, in the case of economic services their production is adjusted to demand, which seems different for ES-services.

positively affected by human activities (in addition to natural incidents)[4].

In this sense ES can be considered flows of value to human societies as a result of the state and quantity of natural capital (TEEB, 2010, p. 7).

The degradation of ecosystems (by human activity) undermine ecosystem functioning and resilience and thus threaten the ability of ecosystems to continuously supply the flow of ecosystem services (de Groot, R., 2012 p.50). Those threats become evident when looking at pollution of air, water or soil, deforestation, climate change and global warming. In this context the question is how to portray degradation of ecosystems. Conceptually ES-degradation is defined as the decline in an ecosystem asset over an accounting period, which is due to economic and other human activities (SEEA-CF, 2012). A question could be raised with regard to how such a change in the ES-asset impacts on the provision of ES-services.

2.2 Scope of ES-services

Regarding the distinction of ecosystems services a categorization into various broad types usually is applied:

- provisioning services, i.e. ES providing air, water, food or timber
- regulating services, i.e. ES providing flood control, pollination or water regulation
- cultural services, i.e. ES offering recreational or cultural benefits.[5]

It has to be mentioned in this context, that there is not a one-to-one relation, but that an ecosystem can provide different services at the same time, a service bundle or basket. As an example, a forest can provide timber (provisioning service), clean the air from carbon (regulating service) and provide the landscape for hiking (cultural service).

An interesting issue concerns the question whether ES services can be negative, i.e. produce so-called dis-services. Looking at nature as a self-regulating and self-balancing system, it seems obvious, that natural processes exist, which result in damages for the welfare of human societies. Practical examples are hurricanes/tycoons, earthquakes, pests etc. Some compare them to negative externalities, which in economic accounts are not covered and hence could be neglected as well (Obst et al., 2016).

2.3 Separate or Holistic Environmental Methodology?

At first glance an astonishing point seems that environmental accounting is split in two separate parts, including separate methodologies, classifications

[4] A difference to economic capital, like machinery, seems to be that the latter has a certain limited service life after which it is useless.

[5] A more detailed overview of what is covered in these categories is provided by table 2 in Costanza et al., 2017, p. 7). A revised version V5.1 of the classification of ecosystem services has been adopted recently (Haines-Young, R., et al. 2017), but the above mentioned three major sections are kept.

and indicators. On one hand the SEEA-2012-CF mainly covers energy, water and material benefits supplied by nature entering the economy as well as certain residuals returned to nature from production and consumption. On the other hand, the SEEA2012-EEA focuses on the non-material benefits of nature. Such a division of methodology is completely acceptable when developing methods and sources as well as testing is the primary goal. Differing priorities at local, regional and national level may be a further reason. But taking into account the decades spent by researchers to work on the different environmental topics, the question could be raised, why not have a single conceptual basis for environmental accounting. The idea is to draft a common, more generic methodology, possibly supplemented by technical handbooks looking at the different ecological areas. As a primary goal such a unified single methodology could strive for elaborating a holistic indicator reflecting the ecological situation in one headline figure. Similar to GDP as a holistic indicator for the economic development, such a single key indicator would make environmental issues much more visible. This in turn should attract the attention of people and broaden the perspective of progress. In addition, more detailed data could be provided for research and for evidence-based decision making.

3. Measurement issues of ES

3.1 Physical Data

To be able to implement the concept of ES-services and ES-assets source data are needed, possibly supplemented by estimations. Such data may for instance be provided by so-called land cover surveys, showing the spatial area by type in square-km as well as the condition of each land type (on a representative basis), to be able to determine the impact of human activities on ES-assets between two reporting dates. A usual classification is to distinguish between sea, coastal areas, land etc.[6]

To estimate ES-services by land coverage type a further step is required, since each land cover type may provide different ES-services at the same time. For instance, a forest may provide timber and wild fruits (provisioning-services) and at the same time hiking opportunities (cultural service). If this is systematically estimated by type of land cover and type of service, the physical flows of ES can be assessed and presented in the following cross-table / matrix (SEEA- EEA, 2012, table 2.2),:

[6] A more refined classification of land coverage types is given by the Land Cover Classification System of the FAO (in SEEA-EEA, 2012, table 2.1):

Type of ecosystem sevices (by CICES)	Type of LCEU				
	Forest tree cover	Agricultural land[a]	Urban and associated developed areas	Open wetlands	…
Provisioning services	For example, tonnes of timber	For example, tonnes of wheat			
Regulating services	For example, tonnes of CO_2 stored/released	For example, tonnes of CO_2 stored/released	For example, tonnes of CO_2 stored/released	For example, tonnes of phosphorus absorbed	
Cultural services	For example, number of visitors and hikers		For example, hectares of parkland	For example, hectares of habitat for ducks	

a Medium to large fields of rain-fed herbaceous cropland.

However, an aggregation of the different ES-services in physical terms is complicated due to the different physical measures used. A proposal mentioned to overcome this difficulty is to use weights for the different ES-services (like in price statistics) and combine them with the physical measures transformed to index numbers. This allows compiling growth rates for total ES-services provided as well as for the basket of ES-services provided by each land cover type.

3.2 Monetary Data

The advantage of compiling data in monetary unit is that a combination with other monetary data is enabled, which may be useful for analytical purposes as well as political discussion. Unfortunately, in practice monetary values for ES-assets and ES-services are missing in most cases. This requires looking for substitutes, i.e. imputed or hypothetical prices.

In economics a value usually consists of two elements, a quantity of a product and the respective market price, i.e. when buyers and sellers are willing to exchange the product. This exchange value guarantees consistency between the records of both parties and therefore is a cornerstone of national accounts. But market prices are not always available, so that in certain cases alternative valuation methods are accepted as proxy. But there are only few exceptions to using market prices as valuation basis for products in national accounts:

- the market prices of similar items is applied for instance in the case of dwellings used by their owners (based on market rents for comparable rented dwellings)
- the costs of production are used for government services, which reflect the market prices for the inputs used, i.e. labor and capital.

Market prices for environmental products can be available sometimes, particularly when these are used as inputs for production, e.g. mineral resources. But also, for certain ES-services market prices do exist, for instance in the case of provisioning services such as crops, fish or water, i.e. products which are consumed by people. Non-consumptive ES-services, like the spiritual or cultural importance of a landscape, are rarely valued in monetary terms (TEEB, 2010, p.7)[7]. So, in the case of ecosystem services researchers need to be innovative to develop a satisfactory valuation. As can be seen from a recent conference paper, 11 different valuation methods are presented including an assessment on their suitability (Obst, C., 2018). However, researchers still have different opinions concerning the fitness of certain ecosystem services valuation methods. This concerns for instance the use of shadow pricing methods, of preference-based methods and of restoration costs (Droste, N. et al 2017). This does not imply that these methods do not provide useful information in certain (regional) circumstances, but it raises questions with regard to the comparability of such valuations over space, nations and time. This leads to the question whether an extension of natural resources assets to cover ecosystem assets is realistic. The main issue is that most economic assets are not actively traded during an accounting period, like dwellings, factories or large customer-tailored software. This holds also in the case of ESassets, so that alternative valuation approaches are needed. One possibility is to compile the net present value (NPV) of future flows of income for each type of ES-asset, which requires mainly - the choice of discount rate - the expected asset life.

If it can be assumed that ecosystem assets tend to exist almost eternally, a realistic estimation of the expected asset life is extremely difficult. In addition, the application of any discount rate for such a long period will heavily influence and probably even dominate the results.

Insofar the following conclusion drawn would seem understandable: “Note that expressing values in monetary units can be a time and resource intensive exercise and often quantitative insights expressed in bio-physical units are sufficient to communicate benefits (e.g. number of people benefitting from clean water provision). Valuation should therefore only be done where it is needed.” (De Groote et al. 2012, last page).

4. Disseminating and combining ecosystem data

4.1 Separate ES-figures

As soon as reliable data on ecosystem services (and assets as well) become available, either in quantity or in value terms, the question of how to publish them has to be addressed. The immediate possibly is to disseminate them as

[7] An exception to this is a recent study to estimate the value of nature-based (short-distance) recreation in Europe for 2012, based on the potential visits of the population for recreational purposes by type and in a regional breakdown. Monetary values are obtained by applying the so-called zonal travel cost method (TCM), based on travel expenses by car (Vallecillo, S. et al., 2019).

stand-alone data and showing their development over time and space. This means either the current amount of ES services provided to human societies or the available stock (and state) of ES-assets (or both).

Such a dissemination strategy could be extended by combining ES-figures with other data and thereby presenting them in a context. An example is to publish data on ES-services (or ES-assets) alongside with GDP development. This would highlight whether economic progress has been achieved on the account of the environment.

4.2 Integration of ES-data and well-being indicators

Various indicators and indicator sets have been developed during the last decades, which differ in certain respects, but having the goal to measure societal progress in a more comprehensive way. At least for indicator sets, like the How's life data set published by the OECD since 2011 (OECD, 2018) to add an indicator on ecosystem services should be possible. But also for other combined indicators, which are based on combining a certain number of factors, like the Human Development Index (UNDP, 2018), the Index of Sustainable Economic Welfare (Nordhaus, W.D., Tobin, J., 1972), or the Genuine Saving Index of the Worldbank (Hamilton, K. et al. 2014), the question could be raised, what is the difficulty to include ES-data as an additional (or replacing) item?[8]

5. Official Statistics on Ecosystems

The fundamental principles of official statistics, endorsed by the General Assembly of the United Nations in March 2014, acknowledge the importance of nature for societies by asking for data on the environmental situation, alongside with figures on the economic, demographic and social situation. "To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information." (UN, 2014, principle 1). Obviously such a test may be based on different criteria. To assess the practical utility in this paper the five criteria developed for statistical products in the European Statistics Code of Practice are used deliberately (Eurostat et al, 2017):

- Relevance

  Relevance means to meet the needs of users. The main reason for this seems that resources are needed for producing official statistics with high quality. But since the users differ, possibly their needs do as well. In the case of ecological data, the general public may be interested in obtaining one or few key-figures on the environmental situation, whereas researchers

[8] This also applies for indicators developed at national level (cf. Durand, M., 2015, annex 1): For Germany for instance: W3-Indicators of the German Federal Parliament (Deutscher Bundestag, 2013), Prosperity Quintet (Denkwerkzukunft, 2014) or KfW-Sustainability Indicator (KfW, 2014)

maybe more interested in detailed data and time series. Furthermore, the needs of stakeholders at the political and government levels are important, so that official statistics should offer inputs to address political agendas both, at the national and international level. At the international level this concerns for instance the UNAgenda 2030 or the Paris Climate Agreement. Against the background of the growing world population, information on the environment with regard to tipping-points would seem particularly useful as well.

- Accuracy and Reliability
  Accuracy may be translated as a good portray of the reality and reliability as giving very similar results when repeatedly compiled. If comprehensive data sources are available for compiling physical data on natural resources and returns to nature there seems no major obstacle that these are produced and disseminated by statistical offices. The situation for value data seems a bit different. On one hand different valuation methods are proposed, for which it is unclear if they produce reliable results. On the other hand in case of some methods their suitability assessments diverge. Insofar it seems that further research is needed. And for statistical offices the advice could be to follow a more cautious approach concerning value data. Possibly they could be published not as official statistics but as socalled experimental data, compiled in close connection and with the support of specialized researchers.
- Timeliness and Punctuality
  Timeliness concerns the interval between the end of a reporting period and the publication date, which should be as short as possible; punctuality concerns the respect of pre-announced publication dates. In digital societies timeliness is very important, because news need to be as topical as possible to attract attention and to serve as a basis for political debates and measures. A much appreciated but challenging timeliness goal consists in publishing environmental data one year after the end of reporting period. This implies to look at the statistical production processes and possibly streamline them for instance by employing satellite images instead of surveys or using company information on non-financial reporting like the standards on sustainability accounting (SASB-standards, 2018).
- Coherence and Comparability
  Comparability of the data refers to the common methodological standards used over a reasonable period of time and across countries; coherence means that arithmetic and accounting identities are respected. Due to the fact that at the moment two separate international methodologies, namely the SEEA-2012-CF and the SEEA-2012-EEA, are used to produce environmental accounts, there is a risk of both double-counting and undercoverage. Therefore, it would seem a useful idea to develop one

(integrated) methodology for environmental accounting. From a dissemination point of view it would seem advantageous to have one single key-indicator being able to portray the ecological situation in one figure.

- Accessibility and clarity
  Clarity refers mainly to the availability of metadata, whereas accessibility concerns the easiness of obtaining data and support. As in other statistical areas, environmental data produced by statistical offices should be published by them via a press release, including a comment (link), so that further information can be obtained. Just disseminating the environmental data is an important first step. In addition, there is a need to produce and publish metadata in order to inform users about methods used including assumptions made and comments on limitations.

For the time being several countries, like the member-states of the European Union or Australia and Canada, regularly produce ecological data, mainly based on the SEEA-2012-CF. For the future relevance of official statistics it would seem useful to work on ecosystem data as well, at least in physical terms. This could be done in a stepwise parallel approach, i.e. working on national priorities and at the same time joining international efforts to make progress in the case of stated international priorities, like the Agenda 2030 or the Paris climate agreement.

6. Conclusion

The focus of this paper is on environmental accounting as well as the ecosystem approach, which are briefly presented on a conceptual basis. A question raised in this respect is, why not have one integrated methodology for environmental accounting instead of two separate ones.

A couple of selected measurement issues are addressed as well, including the issue of compiling ecosystem data in physical and/or in value terms. In addition, different possibilities how to disseminate and combine data on ecosystems are presented, having the aim to attract more attention and broaden the perspective of progress. Finally an assessment on the basis of certain criteria is provided, regarding the question whether official statistics should embark on producing and publishing data on the environment including ecosystems. The recommendation is that by working on ecosystem data statistical offices their relevance is enhancing.

References

1. Brouwer, R., et al, (2013): A synthesis of approaches to assess and value ecosystem services in the EU in the context of TEEB – Final Report http://ec.europa.eu/environment/nature/biodiversity/economics/pdf/EU%20Valuation.pdf

2. Costanza, R., et al (2017): Twenty years of ecosystem services: How far have we come and how far do we still need to go? https://www.robertcostanza.com/wp-content/uploads/2017/02/2017_J_Costanzaet-al.-20yrs.-EcoServices.pdf
3. De Groot, R., (2012) Global estimates of the value of ecosystems and their services in monetary units, Ecosystem Services 1 (2012), 50-61.
4. Denkwerkzukunft (2014), The Prosperity Quintet 2014 - Measuring Prosperity in Germany and Other Early Industrialized Countries, Bonn 2014. http://www.denkwerkzukunft.de/index.php/englishdocuments/index/ProsperityQuintet
5. Diefenbacher, Hans, Zieschank, Roland, et al. (2016): NWI 2.0 –advanced and updated national welfare index for Germany1991-2012 https://www.umweltbundesamt.de/sites/default/files/medien/378/publikationen/texte_29_2016_aktuali sierung_und_methodische_ueberarbeitung_des_nationalen_wohlfahrtsind.pdf
6. Deutscher Bundestag (2013), 17th electoral term, printed paper 17/13300 of 03.05.2013, *Final Report of the Study Commission on Growth, Well-being and Quality of Life – Paths to Sustainable Economic Activity and Social Progress in the Social Market Economy*. An English Summary of the conclusions and main recommendations is available http://www.bpb.de/shop/buecher/schriftenreihe/175745/schlussbericht-der-enquete-kommission
7. Durand, Martine (2015): *The OECD', Better Life Initiative „How's Life" and the Measurement of well-being*, in: International Association for Research in Income and Wealth (edit.), The Review of Income and Wealth, Series 61, no. 1, March 2015.
8. Eurostat et al, 2017: Eurostat and European Statistical System Committee, European Statistics Code of Practice – for the National Statistical Authorities and Eurostat (EU statistical authority), https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db4312-8118-f729c75878c7
9. SNA-2008: European Commission, International Monetary Fund, Organisation for Economic Cooperation and Development, United Nations, Worldbank, System of National accounts 2008, New York 2009.
10. Haines-Young, R., Potschin. M., (2017) Common International Classification of Ecosystem Services (CICES) https://seea.un.org/sites/seea.un.org/files/lg23_cices_v5.1_final_revised_guidance_03-102017.pdf
11. KfW - Kreditanstalt für Wiederaufbau (2014): *Concept Paper – KfW Sustainability indicator* (in German) https://www.kfw.de/PDF/Download-

Center/Konzernthemen/Research/PDF-Dokumente-KfWNachhaltigkeitsindikator/KfW-Nachhaltigkeitsindikator-2014-Konzeptpapier.pdf
12. Meadows, D., (1972): The Limits to Growth;
13. MA [Millennium Ecosystem Assessment] (2005): Ecosystems and Human Well-being: Synthesis. Island Press, Washington, DC.
14. Nordhaus, W. and Tobin, J. (1972): „*Is growth obsolete*?" in: National Bureau of Economic Research, Economic Research: Retrospect and Prospect, Vol 5: Economic Growth
15. Obst, C., et al (2016): Obst, C., Hein, L., Edens, B.: National Accounting and the Valuation of Ecosystem Assets and Their Services, in Environ Resources Econ (2016) 64:1-23
16. Obst, Carl (2018) : The valuation of ecosystem services and assets for SEEA ecosystem accounting, Background paper prepared for BfN/UNSD workshop on ecosystem Valuation, Bonn, Germany-April 2018.
17. OECD (2018): How's Life? 2017 - Measuring Well-being, DOI:https://dx.doi.org/10.1787/how_life2017-en
18. SASB-standards (2018): Sustainability Standards Accounting Board (SASB), Standards for Sustainability Accounting, https://www.sasb.org/standards-overview/download-current-standards/
19. SEEA-CF, 2012: United Nations et al (2014): System of Environmental-Economic Accounting 2012 – Central Framework, New York 2014. https://seea.un.org/sites/seea.un.org/files/seea_cf_final_en.pdf
20. SEEA-EEA, 2012: System of Environmental-Economic Accounting 2012 – Experimental Ecosystem Accounting, New York 2014 https://seea.un.org/sites/seea.un.org/files/seea_eea_final_en_1.pdf
21. SNA-2008: European Commission, International Monetary Fund, Organisation for Economic Cooperation and Development, United Nations, Worldbank, System of National accounts 2008, New York 2009.
22. Sumarga, E., Hein, L., (2014): Mapping ecosystem services for landscape planning, the case of Central Kalimantan, Environ Manag 54 (1): 84-97.
23. TEEB (2010), The Economics of Ecosystems and Biodiversity: Mainstreaming the Economics of Nature: A Synthesis of the Approach, Conclusions and Recommendations of TEEB. http://www.teebweb.org/publication/mainstreaming-the-economics-of-nature-a-synthesis-of-theapproach-conclusions-and-recommendations-of-teeb/
24. UNDP - United Nations Development Programme (2018): *Human Development Indices and Indicators- 2018 Statistical Update*, New York http://hdr.undp.org/sites/default/files/2018_human_development_statistical_update.pdf

25. Vallecillo, S. et al., (2019): Ecosystem services accounts: Valuing the actual flow of nature-based recreation from ecosystems to people; Ecological Modelling, 392 (2019), 196-211;
26. UN, (2014): United Nations, Fundamental Principles of Official Statistics, document A/RES/68/261 of 03 March 2014, https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf
27. UN (2015): United Nations Framework Convention on Climate Change (Paris Climate Agreement) https://unfccc.int/files/meetings/paris_nov_2015/application/pdf/paris_agreement_english_.pdf
28. UN (2016): Transforming our world: the 2030 Agenda for sustainable development, document A/RES/70/1 of 21 October 2015. https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E
29. World Commission on Environment and Development, (1987): Our Common Future / Brundtland Report

Agricultural Census 2020 – How to reduce costs and burden? The European Statistical System approach

Johan Selenius, Christine Wirtz, Denisa Florescu, Andreas C. Lazar
European Commission – Eurostat

Abstract

Information on the structure of agriculture is necessary to understand the risks for ensuring food security, the trends in food production as well as the impacts on rural life and the environment. At global level, the FAO advises all countries to carry out an agricultural census every 10 years to have a full picture of the situation. The European Union (EU) considers it necessary to collect data on farm structure for the common agricultural policy (CAP) and various other policies, often related to environmental aspects, but at more frequent intervals.

This paper presents the technical and methodological aspects of the newly modified system for the European agricultural censuses. Under the new set-up, EU countries can reduce the burden on respondents while increasing the availability of statistics for the census 2020. A set of 184 core variables are collected as a census for all farms above common physical thresholds in all countries with the target of covering 98% of each country's agriculture. For sub-samples of farms, the core variables will be supplemented with variables grouped in three modules on 'Labour force and other gainful activities', 'Rural development' and 'Animal housing and manure management'. The system allows countries to use multiple sources and methods – such as surveys, administrative data or innovative methods and approaches – to obtain data subject to meeting pre-defined quality requirements. The efforts of the European Statistical System (ESS) to allow countries who do not meet the census's 98% agricultural area and livestock coverage requirements to conduct a sample data collection on the smallest farms and further ways of reducing costs and burden will be discussed.

Keywords

Modular approach; Geo-referencing; Eurostat

1. Introduction

Eurostat, the EU's statistical office, publishes EU-wide agricultural statistics on the structure of farms and other operators in the agricultural sector, agricultural production, production factors, and agrimonetary data. Economic accounts for agriculture provide an overall view of the agricultural sector. The main aim is to support decision-making and policy design, implementation, monitoring and evaluation in fields relating to agriculture, such as the EU's common agricultural policy (CAP), sustainable development, and environmental and food-related policies. Challenges in this area of statistics, such as changes in world agriculture driven by globalisation and social change, changes in the CAP and other EU policies, and technological progress and new data sources, have made it necessary to modernise EU agricultural statistics in order to keep providing high-quality, comparable and flexible data (European Commission, 2017). The primary objectives of this ongoing modernisation are to i) meet new and emerging data needs better, more flexibly and faster, ii) improve the comparability and coherence of EU agricultural statistics, and iii) reduce the costs and burdens which data collections place on producers and providers.

The first step was to modify the system for the agricultural census. A new regulation on farm-level structural statistics, Regulation (EU) 2018/1091 on integrated farm statistics (IFS), provides the legal basis for the agricultural census to be conducted in 2020 and two sample data collections to be held in 2023 and 2026. These data collections, which cover the widest range of farms, will update information on the state, trends and impacts of the EU agricultural population. They are thus an essential backbone for other agricultural statistics and also fulfil international requirements and guidelines such as those of the FAO, the Intergovernmental Panel on Climate Change (IPCC) and the United Nations Global Strategy to improve Agricultural and Rural Statistics; for example, the standards, concepts and definitions used in the IFS were designed in accordance with FAO recommendations for better international comparability.

2. Methodology

European statistics are produced by Eurostat in cooperation with the EU member states. The national statistical authorities collect, verify and analyse national data and send them to Eurostat. Eurostat's role is to consolidate these data and ensure they are comparable, using harmonised methodology. The partnership between Eurostat and the national statistical authorities to develop, produce and disseminate European statistics is called the European Statistical System (ESS). In addition to the EU member states, the ESS partnership includes the European Economic Area (EEA) and the European Free Trade Association (EFTA) countries.

National statistical authorities, usually the national statistical institutes (NSIs), collect the data for the 2020 agricultural census. Data collections are based on a modular approach comprising:

- *core* structural information on the most important aspects: land, livestock and the farmer, collected from farms exceeding common physical thresholds. The thresholds were established based on analyses of 2010 census data. The aim is to cover 98% of each EU member state's utilised agricultural area (UAA) and livestock units (LSUs), i.e. economically active farms. For countries where farms above the physical thresholds concerned do not provide the 98% coverage required, further data are collected from a sample of farms below the thresholds. This extends the frame to include small farms (farms in the frame extension, see Figure 1); and

- *modules* with additional information on certain themes which is to be linked to the core (see Figure 1), collected at a lower level of precision and/or frequency from a sample of farms above certain physical thresholds for each module, excluding small farms.

Figure 1: Linking the IFS core and module data collections and frame extension

a. Drivers of cost and burden reduction in the ESS

Under the IFS, the cost and burden of data collection will be reduced by comparison with the previous regulations, Regulation (EC) No 1166/2008 on farm structure surveys (FSS) and Regulation (EU) No 1337/2011 on permanent crops. This will be done in a number of ways, which are labelled with R codes in this section. Under the modular approach, the costs and burden of the agricultural census to be held in 2020 are naturally linked with the sample data collections in 2023 and 2026 and therefore assessed together with them, as any additional efforts in the first reference year (2020) will pay off in subsequent reference years (2023 and 2026).

i. Thresholds and frame extension

The previous regulation on farm structure statistics required all member states to cover holdings above certain physical thresholds. In a few cases, these thresholds were too high to achieve sufficient coverage of UAA and LSUs, and some EU countries collected data from almost all their farms. IFS sets lower physical thresholds, but if the requirement of achieving 98% UAA and LSU coverage is fulfilled, EU countries can apply to raise the physical thresholds, so as to reduce the number of holdings from which data are collected. So the IFS thresholds are flexible and can even be raised above the ones set in the previous regulation if the 98% coverage requirement is met (R1).

Under the IFS, if countries cannot achieve the required 98% coverage when applying the thresholds, they should extend the frame to include a sample of small farms that fall below these thresholds (see Figure 1); this sample should collect only core variables (not all variables, as was the case under the previous FSS regulation) (R2). The frame will be extended to include small farms in 2020 only (not in 2023 and 2026). This contrasts with the previous regulation, which required small farms to be surveyed in each reference year (R3). In addition, the frame extension can be applied to a sample in 2020, whereas a full census was required in 2010, under the previous regulation (R4). The smallest farms are surveyed only once a decade because they produce mainly for themselves; it is assumed that they are fairly homogeneous and change relatively little over time. Their importance lies mainly in the social dimensions of agriculture and rural areas. This change will be beneficial to countries with many small farms, such as Bulgaria and Romania. It represents a departure from the previous system, under which all farms were included in the populations covered by the former censuses and sample data collections.

Under the IFS, modules are generally collected on holdings above given physical thresholds (main frame, see Figure 1). However, the relevant population is further reduced for particular modules (R5). In 2020 and 2026, for example, the 'Animal housing and manure management' module will only be collected from farms with livestock, which account for between 20% and 80% of all farms, depending on the EU country concerned.

ii. Source agnosticism

EU national authorities can use a variety of methods and sources to collect data, provided that they meet the necessary quality requirements. These methods and sources are:

S1: statistical surveys: census (a) or sample (b),

S2: administrative data sources, or

S3: other methods and innovative approaches.

For the census year, the IFS regulation allows samples to a greater extent than the previous regulation. While most variables were collected by census in 2010, the IFS regulation states that a sample-based approach can be used for all modules' variables in 2020 (R6). Censuses and sample data collections can be implemented through surveys (S1a and S1b), by making use of administrative registers (S2), or through other approaches (S3).

Administrative sources (S2) based on EU regulations and where a certain control is implemented can be used without any justification. The previous regulation explicitly mentioned fewer such administrative sources and required prior information on the methods and the quality of the data from additional administrative sources. The IFS explicitly mentions more administrative sources, thereby reducing the number of cases in which prior information on their use and quality is required (R7). The administrative sources which are regulated to support the CAP are the organic farming registers, the vineyard register, the integrated administration and control system, the systems for the identification and registration of bovine and ovine animals and the administrative sources associated with specific rural development measures. For example, in the 2020 agricultural census, the data for the 'Rural development' module on farms benefiting from rural development measures are expected to be collected entirely from administrative registers, while the variables of livestock in the core are expected to be collected from the animal registers.

Other methods and innovative approaches (S3) refer to modelling, expert estimates, remote sensing, and so on. Some of the variables in the 'Animal housing and manure management' module can be estimated using models. For example, the annual average number of animals in each category can be estimated using models based either on the number of animals raised, divided by the number of livestock raising cycles per year, or on a combination of the number of places and the number of empty days. To give another example, the quantity of manure produced by a given category of livestock can be estimated on the basis of the number of animals under certain types of management. Such methods reduce the administrative burden on national authorities and the costs they incur, but also the effort that farmers have to make to recall and estimate this information (R8).

Under the IFS, in order to allow member states to flexibly choose the source and reduce the burden of data collection further, information on the 'Machinery and equipment' (in 2023), 'Orchard' (in 2023) and 'Vineyard' (in 2026) modules may be based on the year directly preceding or following the reference year, as long as it reflects the situation in the reference year (R9). This information is assumed to be slow-changing. Previously, all data had to be collected for the reference year.

While the increased use of S2 and S3 reduces the overall burden, it poses challenges with respect to assessment of data quality. Data quality is affected by both sampling and non-sampling errors. Only data collections based on samples are affected by sampling errors. In this case, the theory for probability samples allows for the sampling design to control for this type of error, and the IFS regulation sets precision targets for certain variables collected on a sampling basis. However, it is impossible to control for non-sampling errors or to target them in advance, because they happen in a non-controlled manner for every statistical process. They cancel each other out or add to or multiply each other, depending on the specific data collection and the specific context. No sound theory is available to predict them. This is more problematic in cases S1a, S2 and S3, where non-sampling errors have a big impact on overall data quality. After the data collections are carried out, multiple indicators can be calculated for each source of non-sampling errors associated with the different steps in a statistical process, while a unique synthetic indicator, such as the mean square error, is not computable. Thus, synthetic quality indicators cannot be set in a regulation such as the IFS. However, national authorities are required to describe in quality reports a set of indicators that are both quantitative and qualitative. Quality assessment for multisource data is under continuous development in ESSnet projects[1] such as the ESSnet on Quality of Multisource Statistics.

The move towards source agnosticism improves subsidiarity, enabling EU member states to choose the most cost-effective sources for themselves according to their national conditions and needs. It also increases flexibility, thanks to greater openness to future data sources such as precision farming data.

iii. Reduced number of variable breakdowns

The 2010 agricultural census covered a total of 273 variables, collected from all EU farms. In the 2020 census, the core will comprise 184 variables and will be supplemented by 30 variables in the module ‘Labour force and other gainful activities’, 15 in the module ‘Rural development’, and 70 in the module ‘Animal housing and manure management’, totalling 299. These are maximum numbers, as member states may transmit fewer variables if particular items do not exist or are not significant in the country concerned.

The above information refers to the absolute numbers of variables in 2020. Burden reduction should be assessed throughout 2020-2026, as the data for the modules will be collected less frequently in the new decade (R10). Moreover, much burden reduction will be achieved in 2023 and 2026 by taking the previous regulation on permanent crops into account as well: in 2023, the IFS will collect a maximum of 470 variables, including the ‘Orchard’ module, whose predecessor regulation contains over 650 variables, while in 2026 a maximum of 350 variables are allowed, including the ‘Vineyard’ module, whose predecessor regulation contains almost 900 variables (R11). In addition, the core and module system enables that fewer farms will have to provide data on all variables in a census or sample year. This reduces the burden on individual farms (R12).

Furthermore, the same list of variables and definitions, as well as common quality standards and data transmission deadlines, will be used in all EU member states, increasing interoperability and reusability so as to reduce the costs and burdens of data collection (R13).

---

[1] An ESSnet project is a network of several ESS organisations which aims to provide results that will be beneficial to the whole ESS.

b. Enhanced data visualisation through geo-referencing

While the IFS regulation allows costs and burdens to be reduced, it enhances the potential for geographical data analysis. The regulation requires information on the geographical location of each farm, namely the code of the cell where each farm is located (using the 1 km INSPIRE statistical units grid for pan-European usage[2]). INSPIRE is based on the infrastructure for spatial information established and operated by the EU member states. It enables environmental spatial information to be shared among public sector organisations and assists in policy-making across boundaries.

Geo-referencing of statistical information at the farm level means that each holding included in a dataset is assigned a high-accuracy geo-reference code. The 1 km grid should be used as a geo-referencing framework. When studying socioeconomic and environmental phenomena such as flooding, a system of grids with equal-size grid cells has many advantages: it allows for easy comparisons, ensures stability over time and allows for hierarchical aggregation, depending on the specific purpose.

INSPIRE provides recommendations on how to assign stable identifiers (codes) to statistical units (agricultural holdings). Having holdings assigned to detailed geographic areas (1 km grid cells) enriches data analysis and allows tabulations to be produced for additional geographic entities that are more detailed than NUTS and cross-border tabulations when quality is good enough, thereby meeting relevant policy needs. For example, data can describe geographic entities such as river basin districts and Natura 2000 sites. The Commission uses hydrological definitions of basins and sub-basins to assess water balance by the basin of main rivers and tributaries in the EU. IFS data is crucial in making estimates like this. Tabular data are treated for confidentiality and suppressed or aggregated to the upper nested grid level[3]. This occurs when tabular data are unreliable, based on the values of estimated sampling errors, or when the 1 km grid contains 10 or fewer agricultural holdings or is too small for the average farm size in the grid. The upper nested grid cells of 5 km, 10 km or even larger are used as required. Reliable data can be displayed using maps for enhanced data visualisation.

3. Results

The new IFS regulation entered into force only in August 2018. It is thus still too early to quantify exactly how much the new system will reduce the burden and costs which data collections entail.

Before developing and adopting the IFS regulation, the Commission conducted a formal impact assessment. On the basis of model scenarios, it estimated that the IFS Regulation would bring net monetary savings of around

[2] Commission Regulation (EU) No 1089/2010

10% of current total costs (almost € 320 million reported in the last agricultural census, conducted in 2010). The following had already been taken into account: external factors (such as the decreasing number of farms in the EU and the technological and other progress expected), adaptation costs (such as redesigning data collection systems and improving coherence and harmonisation), and increased data collection costs (owing to some new variables and earlier deadlines). The expected sizeable drop in the number of farms from which data is to be collected (from 12 million in 2010 to 8.4 million in 2020), owing to the new thresholds and the frame extension, was identified as the main source of these savings. However, the impact assessment could not take account of factors such as the smaller number of variable breakdowns and the effects of greater source agnosticism, beyond general estimates, as they had not yet been finalised at the time (European Commission 2016). This means that while the 2020-2026 round of census and sample data collections can be expected to cost less and impose fewer burdens than the 2010-2016 round, no exact numbers will be available until later on in this decade.

4. Discussion and Conclusion

Organising the collections of data under the IFS regulation is a challenge for all ESS participants. After the 2020 agricultural census, which – given the innovations and efforts outlined in this paper – is expected to account for most adaptation costs, the costs and burden are almost certain to shrink further. Moreover, the system will probably be able to adapt more easily and thus at a lower cost to changing and emerging needs.

Agricultural, rural and related policies account for significant expenditure under the EU budget, coming to around € 60 billion per year, which is about 40% of the total EU budget. They create jobs and promote sustainable growth in the EU, and their impact on the environment, food security and rural communities is very considerable. Developing, implementing and monitoring these policies requires an evidence basis of high-quality, up-to-date official statistics. The cost of the 2010 agricultural census totalled about € 320 million across the EU and its member states. Official statistics are thus a comparatively cheap public service and an investment that generally pays for itself, but high costs and heavy burdens can nevertheless jeopardise their production and acceptance at a time of tightening public finances. This paper has presented the ESS's way of tackling this challenge, which we trust will be instructive.

[3] INSPIRE Data Specification on Statistical Units, Directive 2007/2/EC, Annex

References

1. Regulation (EU) 2018/1091 of the European Parliament and of the Council of 18 July 2018 on integrated farm statistics and repealing Regulations (EC) No 1166/2008 and (EU) No 1337/2011, OJ L 200, 7.8.2018, pp. 1-29. https://eur-lex.europa.eu/eli/reg/2018/1091/oj.
2. Commission Implementing Regulation (EU) 2018/1874 of 29 November 2018 on the data to be provided for 2020 under Regulation (EU) 2018/1091 of the European Parliament and of the Council on integrated farm statistics and repealing Regulations (EC) No 1166/2008 and (EU) No 1337/2011, as regards the list of variables and their description, OJ L 306, 30.11.2018,pp.14-49.https://eurlex.europa.eu/eli/reg_impl/2018/1874/oj.
3. Regulation (EC) No 1166/2008 of the European Parliament and of the Council of 19 November 2008 on farm structure surveys and the survey on agricultural production methods and repealing Council Regulation (EEC) No 571/88, OJ L 321, 1.12.2008, pp. 14-34. https://eur-lex.europa.eu/eli/reg/2008/1166/oj.
4. Regulation (EU) No 1337/2011 of the European Parliament and of the Council of 13 December 2011 concerning European statistics on permanent crops and repealing Council Regulation (EEC) No 357/79 and Directive 2001/109/EC of the European Parliament and of the Council, OJ L347,30.12.2011,pp. 7-20. https://eur-lex.europa.eu/eli/reg/2011/1337/oj.
5. European Commission (2016): Commission Staff Working Document Impact Assessment accompanying the document Strategy for Agricultural Statistics 2020 and beyond and subsequent potential legislative scenarios, SWD/2016/0430.https://eur-lex.europa.eu/legal- content/EN/ALL/?uri=CELEX:52016SC0430.
6. European Commission (2017): Commission Staff Working Document Evaluation accompanying the document Strategy for Agricultural Statistics 2020 and beyond and subsequent potential legislative scenarios, SWD(2017)96/F1. http://ec.europa.eu/transparency/regdoc/rep/10102/2017/EN/SWD-2017-96-F1-EN-MAIN-PART-1.PDF.
7. Commission Regulation (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services. https://eur-lex.europa.eu/eli/reg/2010/1089 .
8. INSPIRE Data Specification on Statistical Units, Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services, Annex 3, https://inspire.ec.europa.eu/id/document/tg/su .

9. ESSnet on Quality of Multisource Statistics. https://ec.europa.eu/eurostat/cros/content/essnet-quality multisource-statistics-komuso en .

STATLEARN@: An e-Learning tools for assisting students' understanding of hypothesis testing

Nor Azila Che Musa[1], Zamalia Mahmud[1], Norhayati Baharun[2], Azizan Haniff Adenan[1]
[1]Universiti Teknologi MARA Shah Alam
[2]Universiti Teknologi MARA Cawangan Perak

Abstract

In an advanced world of technology and communication today, the use of technology in teaching and learning is becoming more important and significant in various academic programmes. Educators and instructional technologists are developing various electronic teaching and learning tools that enable us to capture the attention of students as well as to complement the traditional teaching method. In the teaching of Statistics for Business and Social Science subject at the Universiti Teknologi MARA, instructors are faced with a greater challenge as statistics is not the core subject in their respective programme of study. Hence the lack of knowledge in statistics and statistical software among the students are obvious. However, the current generation of students are more comfortable in handling multimedia technologies, hence this led to the development of STATLEARN@, an e-learning tools to assist or support them to learn statistics more effectively. STATLEARN@ is developed following the student's difficulty in understanding the concept of Hypothesis Testing. After taking into account how students should understand the process of conducting hypothesis testing, an online open-source platform was created. This include the development of e-learning contents/notes, quizzes and tutorials which are appropriate for the respective topics. However, learning statistics through the e-learning platform is one of the biggest challenges for students. Since statistics has to deal with calculation and computation as well as grasping the theory behind it, students need a proper guide on how to solve the statistics problems through the computer dashboard. Hence in STATLEARN@, the learning process is enhanced through a systematic guide where interactive contents are placed when needed by the learners. STATLEARN@ has become an attractive tool for students with slow learning capability. This paper will describe how STATLEARN@ is used to assist students in learning and understanding about hypothesis testing in a relax environment. STATLEARN@ is designed to help meet the needs of learners and at the same time assists learners to probe for correct answers with the support of the interactive notes. STATLEARN@ also provide students with a step by step practice guide and test questions to help students increase their understanding on hypothesis testing and practice their statistical skills in solving real problems. Based on the outcome of a pilot study, it shows that students who have used STATLEARN@ discover a better learning experience that can improve their understanding of statistics.

Keywords
e-learning; hypothesis testing; statistical problem solving; understanding

1. Introduction

The advancement of technology and existence of internet has led to the expansion of innovative teaching and learning methods in higher education institution such as virtual and electronic learning. E-learning is defined as distance learning by using electronic media as medium (Kaewkiriya, 2013) and as acquisition of knowledge and skill using electronic technologies with internet-based networks (Thinh, 2016). E-Learning is basically a web-based system that makes information or knowledge available to users or learners and due to the opportunities created by e-learning, it bridges the gap between learners and instructor at anytime and anywhere regardless of their geographical locations (Sun, Tsai, Glenn, Chen & Yeh, 2006). In other words, e-learning employs learners in the learning process, provides occasions for learners' online activities at their own pace and at the time suitable for them. It creates learning environment and makes them responsible for their own learning. Moreover, e-learning improves learners' time-management skills, ability to organize their study time and capable of prioritizing which part to focus on (Songkram et. al., 2015; Oproiu, 2015; Nugroho, 2016).

The learning management system (LMS) is widely used nowadays. It helps to convert physical medium to virtual server which contain the knowledge and information (Mohd Shahizan et. al, 2012). LMS provides the instructor to create and deliver content, monitor student participation, and assess student performance with interactive features (Ghoniem, Aljahdali, & Fahmy, 2010). Nowadays, e-learning has become one of the main tools and widely used in higher education institutions, hence the changes and development should be made in the form of managing, creating and developing e-learning platform which suit the needs of users (Nakorn, 2004). Suitable learning materials and contents are very important things to consider when developing e-learning platform because it will give a big impact to the effectiveness of learning (Kaewkiriya, 2013).

There are many technologies that have been developed and accessible for statisticians to utilize in their teaching. E-learning platforms are also developed to facilitate teaching and learning. Universiti Teknologi MARA is active in promoting the use of i-learn platform in an effort to motivate lecturers to teach using electronic resources. However, the use of e-learning in teaching statistics at the Malaysian higher education institutions is still in the early stage as most instructors and statistics educators are still grappling with the latest technology to create e-learning materials.

It is acknowledged that instructors are faced with various challenges in teaching statistics to students with no prior statistical knowledge. Students are more comfortable with multimedia technologies and internet to assist them to learn statistics more efficiently. Therefore, this paper will particularly focus on the development of STATLEARN@, an electronic platform and resources in facilitating students' learning of Statistics for Business and Social Sciences subject.

2. About STATLEARN@

Developing an e-learning platform that contains interesting features of information and user-friendly interface literally involve complex process. There are many methodologies that are different or even similar that have been developed to help in building the e-learning management system. A developing method is observed as a procedure or path by the developer that proceeds from a problem of a certain class to a solution of a certain class (Jackson, 1982).

STATLEARN@ is developed for the purpose of assisting the non-statistics major students to enhance their understanding of statistical concepts particularly in hypothesis testing. It is observed that many students experience some difficulties in understanding the concept of hypothesis testing. After taking into account how students should understand the process of conducting hypothesis testing, an online open-source platform was created. This include the development of e-learning contents/notes, quizzes and tutorials which are appropriate for the hypothesis testing topics which include 2-independent sample t-test, 2-dependent sample t-test and one-way ANOVA. However, learning statistics through the e-learning platform can be a challenge for any students because they are used to the instructor's face-to-face teaching in class. The development involves the process of making student understand what statistical concepts need to be learned and how it can be applied to solve problems using real data. In view of this, students need a proper guide on how to solve the statistics problem through the computer dashboard. Hence in STATLEARN@, the learning process is enhanced through a systematic guide where interactive contents are placed as needed by the learners.

There are three main sections in STATLEARN@, which are notes and formula in section 1, step by step practice questions in section 2 and test questions in section 3. Step by step practice questions enable the students to solve a hypothesis testing problem interactively based on the prompts given at each step. The practice section requires students to solve the problems by providing answer at each step correctly before he/she can proceed to the next step. The system will capture the amount of time taken (in seconds) spent at each step. In the test questions section, students are given specific time to attempt the test and their performance measured at the end of the test.

3. Development of STATLEARN@

STATLEARN@ is developed as a web application. The major reason why web application was chosen as the platform for STATLEARN@ is because it improves the accessibility and the availability of STATLEARN@, which in turn allows students to access the STATLEARN@ anytime at any place, provided they are connected to the internet. STATLEARN@ was developed using basic web development languages. HTML, CSS and JavaScript are used for the front-end design, in other words, the presentation, of the application. A popular CSS Framework named Bootstrap is used to achieve responsive web design. With responsive web design, STATLEARN@ becomes a mobile-friendly web application, allowing students to use the application easily on their smart phone devices. Moreover, a popular JavaScript library named jQuery is used to implement interactivity on the front-end side of the application. For the back-end side of the application, PHP is used. The back-end consists the logic of the application, for example the functions to check for students' answers as well as the functions to perform calculations. This development process will calibrate with the process of solving hypothesis testing. It can be explained by the flow chart in Figure 1.

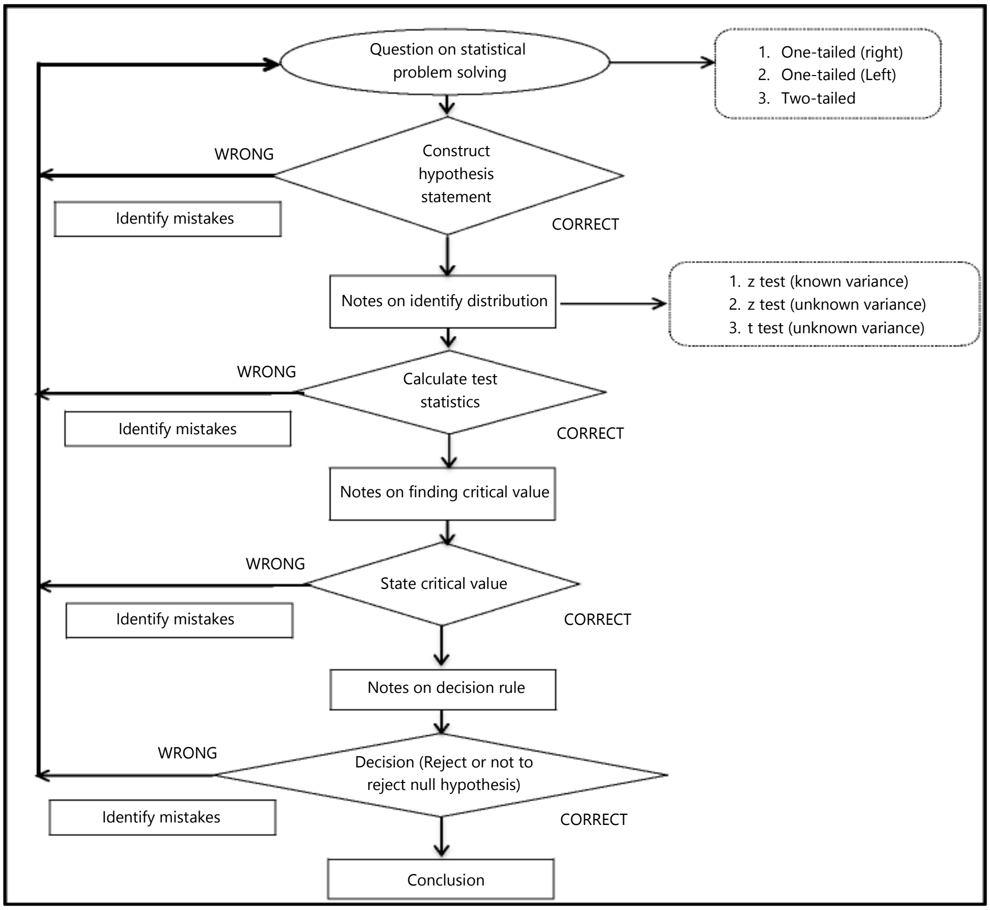
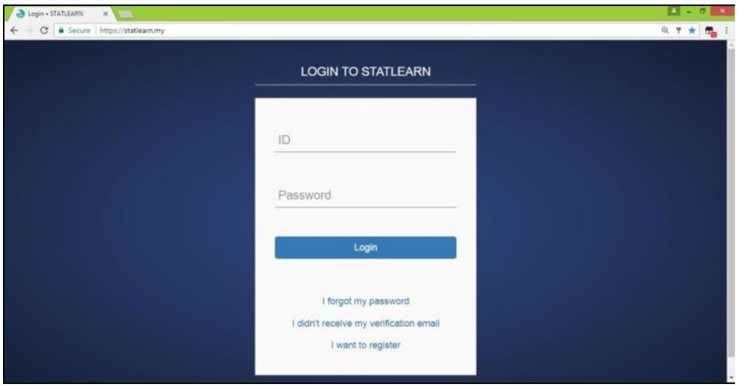


Figure 1: Flow chart of hypothesis testing process using step by step practice questions

4. STATLEARN@ Interface

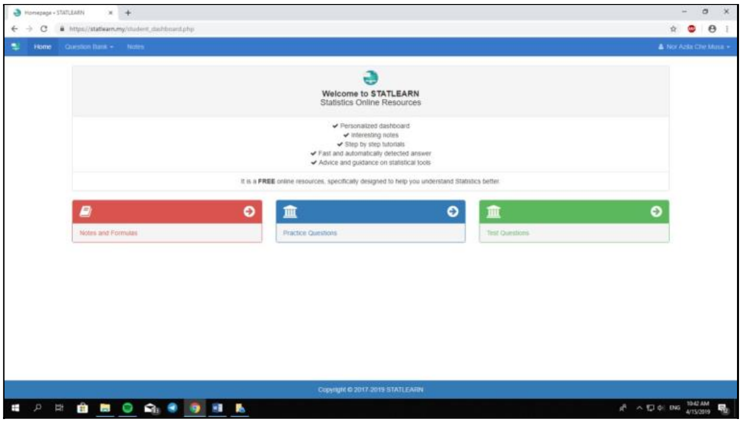
The following illustrates the interface for STATLEARN@. Figure 2 shows the first/homepage which require student to login their username and password.



Figures 2 : First page of statlearn.my

Students are required to register on the website by filling in their personal information. Upon receiving email verification from the administration, students are allowed to login to the website using their student ID and password.

Figure 3 shows the main page of the website once students has successfully logged in. There are three main menus: (1) notes and formula, (2) practice questions and, (3) test questions.



Figures 3: Main page of statlearn.my

The practice questions section contain ten questions. Students may choose which questions they want to attempt regardless of the sequence. While doing the questions, students can refer to the notes and formula provided at the bottom of every page. Figure 4 shows the interface of the questions for Step 1.

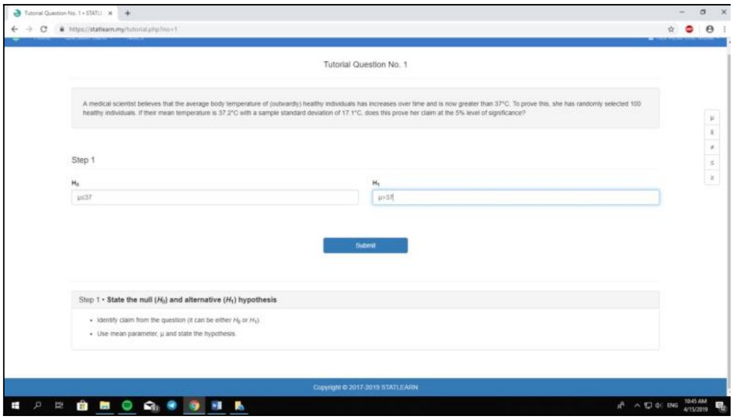


Figure 4: First page of Question 1

STATLEARN@ also provide questions for students to test their understanding. There are six questions for students to answer within the duration of three hours. Students have the flexibility to logout and login at any time they wish. The timing clock will pause once students has logged out and the remaining time will be displayed at the top of the questions, as shown in Figure 5.

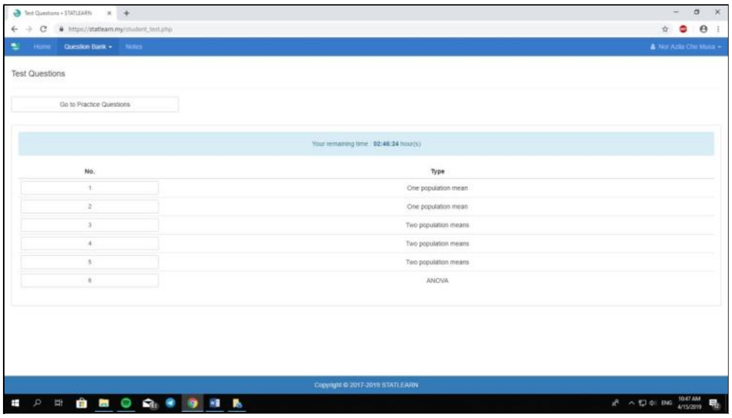


Figure 5: List of Test Questions

5. Conclusion

Based on the feedback that we received from the students, e-learning is a favourable alternative medium to the traditional classroom learning. They discovered that STATLEARN@ can enhance their understanding about hypothesis testing. They also find that STATLEARN@ has the interface and features that enable them to learn and practice doing statistics exercises easily with the help of interactive built-in materials. STATLEARN@ also provides flexibility and convenience in assisting students to learn statistics using any mobile device regardless of their locations. It is particularly useful for slow learners and definitely complement classroom teaching. E-learning will keep growing as an essential part of academic and professional education. However, more effort are needed to include the right materials and create a more appealing and effective online learning environment (Goyal, 2012).

References

1. Ghoniem, S., Aljahdali, S., &Fahmy, A. (2010). A dedicated Web-Based Learning System. Universal Journal of Computer Science and Engineering Technology, 84-92.
2. Goyal S, (2012). E-Learning: Future of Education, Journal of Education and Learning. Vol.6 (2) pp. 239-242.
3. Jackson, M. (1982). A System Development Method. In Tools and Nation for Program Construction: An Advanced. Cambridge: CUP.
4. Kaewkiriya, T. (2013). A Design and Development of E-Learning Content Ffor Multimedia Technology using Multimedia Game. International Journal of Software Engineering & Applications (IJSEA), Vol.4, No.6.
5. MohdShahizan, O, Nadirah, M, Lizawati, M. Y., Norazah, Y., and Shaffika, M. S.,(2012). An analysis of e-learning system features in supporting the true e-learning 2.0, ICTLHE ● RCEE ● RHED

6. Nakorn, T. (2004) Design e-Learning concept,Ramkumhang Journal , BKK Thailand, Vol. 3 No. 3, pp. 143-145.
7. Nugroho, L. E. (2016) iJET, 11(01), 62–65.
8. Oproiu, G. C. (2015) Procedia Social Behaviour Science,180, 426– 432.
9. Prawira, A. K., Sofianti, T., D., Indrayadi, Y., (2015). Developing E-Learning System to Support Teaching and Learning Activities Using DSDM Approach, Performa Vol. 14, No.1: 41-52.
10. Thinh, D., V. (2016). The Role of E-learning. Management, Enterprise and Benchmarking in the 21st Century, Budapest.
11. Songkram, N.,Khlaisang, J., Puthaseranee, B., Likhitdamrongkiat, M., (2015) Procedia Social Behaviour Science, 174, 667–673.
12. Sun, P. C., Tsai, R. J., Glenn, F., Chen, Y. Y., &Yeh, D. (2006). What drives a successful eLearning? 2-3.

Data skills in schools: Teaching Statistics

David Young[1,2], John Reilly[3]
[1]University of Strathclyde, Glasgow, UK
[2] NHS Greater Glasgow and Clyde, Glasgow, UK
[3]Education Scotland, Glasgow, UK

Abstract
In Scotland, numeracy across learning has been emphasised as a key area of the Curriculum for Excellence to provide learners with essential analytic, problem-solving and decision-making skills. The Statistics Award was developed in collaboration with the Scottish Qualifications Authority and the University of Strathclyde in an attempt to provide an opportunity for young people to learn these skills. The emphasis of the course is on the application of statistics to real-life data and interpretation of results in an appropriate context. Minitab or R Studio are used for the analysis of data with learning skills focussed on the application of appropriate statistical methodology and clear interpretation of results in the context of the research question of interest. Support materials for teachers and students have been developed including online resources funded by the Nuffield Foundation's Q-Step programme.

Keywords
Teaching; Teaching statistics; Statistical literacy; Statistical software;

1. Introduction

Many organisations are struggling to recruit staff with skills in data science. It is thought that Scotland needs about 13,000 additional workers each year to cope with the demand for data skills [1]. Two of Scotland's universities have launched a drive to train 100,000 Scots in data skills over the next decade in an effort to transform the country into the 'data capital of Europe'.

The Royal Statistical Society (RSS) has called on the UK government to recognise the value of data and the importance of having the capability to use it [2]. The growth in data collected by technology will play a huge part in the economy of our country and in our personal lives. Investment in education and training to support industry growth is the only way to tackle the gap in skills and preparation to make use of the information stored in data needs to start in early school.

The Making Maths Count Group was set up by the Scottish Government to consider ways of improving confidence and fluency in maths and numeracy for children, young people, parents and those who deliver maths education, to raise attainment and achievement across learning. The Group published a report in September 2016 containing ten recommendations [3]. Recommendation number 5 states that 'Education Scotland should evaluate the quality of children's and young people's learning in attainment in maths and share examples of good practice.' Around forty secondary, primary and early learning and childcare settings are currently being inspected in response to this recommendation and a final report is due to be published in August 2019. Recommendation number 8 focuses upon improving maths skills for employment including understanding data and statistics.

Education in school forms the foundation for further education and training. Employers have an unmet demand for people with strong literacy and numeracy skills that can be applied in a variety of problem solving contexts, and there is a substantial skills gap in data-driven industries. Work in the data economy is strengthened by successful training in statistics and data analytics combined with skills in critical thinking and the communication of findings. Training and professional development for teachers will be a key ingredient for success. In response to meeting these challenges, 'Developing the Young Workforce' (DYW) is a seven year programme that aims to better prepare children and young people for the world of work. Working with partners there is now an expectation that all schools provide flexible pathways meeting the needs of all learners and ensuring positive destinations for all young people. The Higher Statistics Award has been designed to support the aims of this programme by providing a stand-alone qualification at Higher level that can be delivered by staff working in a variety of curricular areas, thereby offering young people an alternative curricular pathway and providing skills for learning, life and work.

The aim of teaching statistics at school level is to develop statistically literate young adults who make reasonable assumptions when presented with a set of data. Too often it is the case that while students develop adequate procedural knowledge of how to compute descriptive statistics and construct graphs, they are less able to reason about where and when to use these [4]. The RSS has stated 'We would like to see basic numeracy and statistical literacy included in teacher training and in continuing professional development (CPD) for the whole profession' [5]. They see this training of teachers from primary school through to university lecturers as a key ingredient for success in encouraging data literacy in young people. They further emphasise the need to ensure that the training is focussed less on mathematical theory and more on the application of practical skills to real-life situations. CPD training for teachers should be applicable since management of data is a major part of their working lives. [6, 7].

This paper describes the development of the Scottish Qualifications Authority (SQA) Statistics Award in an effort to address these needs.

2. Methodology

The SQA Higher Statistics Award [8] was the result of meetings with Deans of Science and Engineering in Scotland which highlighted a lack of data analysis and problem solving skills in higher education. Further consultations were done with several Higher Education institutions to ensure the course content and learning outcomes would be appropriate and relevant and a working group was tasked to generate ideas.

Meetings with various stakeholders i.e. schools, colleges, universities and employers, helped to identify the key statistical skills to be included in the course curriculum. It was clear that the course should be light on mathematical theory with the emphasis being on the application of data science skills. In addition, the course should be available to pupils of all ability levels across the school curriculum.

Course content was agreed and resources and materials designed. These included support notes, course specification and assessments, teaching and learning resources, access to real life data and computer software training and support. Once developed, the award was promoted at CFE implementation events and CPD training for teaching staff was offered at the Universities of Strathclyde and Edinburgh. These events highlighted a need for additional resources, training and support which lead to further development and revision of the course.

3. Results

The overall aim of the Higher Statistics Award is to develop knowledge, skills and understanding in statistical methods and techniques applied to a variety of real-life contexts from across the curriculum, some of which may be new to the learner. Candidates who complete this qualification will be able to:

- use and apply statistical skills in real-life contexts
- identify and perform an appropriate statistical analysis on given data set(s) using a statistical software package
- communicate the results of a statistical analysis, clearly and concisely, in the context of the problem being addressed

3.1 Course content

Content was devised in collaboration with industry and education to ensure fundamental data science skills covering a wide range of applications would be covered within the course. There are three main topics: an introduction to statistics, correlation and regression and hypothesis testing.

1. Introduction to Statistics: includes types of data; summarising data graphically and numerically using appropriate descriptive statistics; sampling and data distributions (specifically the normal distribution)

2. Correlation and Regression: the simple mathematics of a correlation coefficient is described in detail to ensure a good understanding of the concept and specifically when the use of correlation is inappropriate; mathematics underlying regression modelling is covered lightly, with no need of an understanding of differentiation; model predictions and prediction intervals are explained and goodness-of-fit is assessed using the coefficient of determination; passing mention is made of multiple regression and analysis of residuals
3. Hypothesis Testing: basic ideas; null and alternative hypotheses; p-values; Type 1 and Type 2 errors; confidence intervals; study design; tests for paired and independent samples

3.2 Approaches to learning and teaching

Effective learning and teaching draws on a variety of approaches to enrich the experience of learners. In the Statistics Award, a mix of approaches are used which provide opportunities for personalisation and choice to help motivate and challenge learners. Some of these approaches include: interdisciplinary learning, cross-curricular approaches, investigative and problem solving approaches and resource based learning and e-learning.

To achieve this aim, learners will engage in a variety of learning approaches and activities as appropriate to the understanding of statistical concepts, for example:

- using active and open-ended learning activities such as scientific research, case studies, project-based tasks and presenting findings to others.
- using real-life contexts and experiences that are familiar, unfamiliar and relevant to young people, to meaningfully hone and exemplify skills, knowledge and understanding
- making use of the internet to draw conclusions about specific issues
- recording, in a systematic way, the results of an investigation from different sources
- participating in group work with peers and using collaborative learning opportunities to develop team-working skills
- develop problem solving and critical thinking
- use of questioning and discussion to engage learners in explaining their thinking and check their understanding of fundamental concepts
- making links in themes which cut across the curriculum to encourage transferability of skills, knowledge and understanding — including technology, geography, sciences, social subjects, mathematics, applications of mathematics and health and wellbeing
- using written and/or oral communication and presentation skills to present information

- using appropriate technological resources (e.g. web-based resources)
- using appropriate media resources (e.g. video clips)

Some of these learning and teaching activities may be carried out on a group basis and, where this applies, learners can also receive feedback from their peers. Where possible, teachers and lecturers should provide opportunities to personalise learning and enable learners to have choices in approaches to learning and teaching. There may be opportunities to contextualise approaches through mini-projects or case studies.

Teachers and lecturers should also create opportunities for, and use inclusive approaches to learning and teaching. This can be achieved by encouraging the use of a variety of strategies suited to the needs of all learners. Innovative and creative ways of using technology are valuable in creating inclusive learning and teaching approaches.

3.3 Use of Statistical Software

In order for learners to be equipped with skills in practical data analysis, it was widely agreed that use of a statistical software package should be a key part of the course and, for ease of use, Minitab [9] was the initial software of choice. The introduction of a practical element to the course ensures that candidates are skilled in applying statistical methods to real world problems and able to manipulate and analyse large data sets. The use of a propriety statistical package within the teaching framework also sets this course apart from the Advanced Higher Statistics course offered by the SQA which is more focussed on probability theory and the mathematical aspects of statistics and does not have a practical data analysis component [10].

The key learning outcomes require candidates to be proficient in using a statistical software package for data manipulation and analysis. The licence cost associated with Minitab proved to be prohibitive for schools with limited budgets and R Studio [11] was considered for the practical part of the course. The main advantage of R Studio is that it is a free, open source statistical analysis package and therefore makes the course accessible to all schools in Scotland. The disadvantage is the lack of ease of use. The use of R Studio for the statistics unit was piloted on a group of local school children, aged 15-17, at the University of Strathclyde. The candidates were already using Minitab and subsequently used R Studio to tackle the same research questions. There was no clear preference among those candidates for Minitab over R. In order to ensure that the use of R Studio would not be a barrier to learning, and also to help in training for teaching staff, an online resource was developed for full implementation of the course using R Studio. The development of this resource was funded by Q-Step and two interns – both undergraduate psychology students – produced a web-site with complete teaching and learning resources using R Studio [12].

3.4 CPD Provision

Since the RSS has advocated CPD for teaching staff so strongly, it was important to provide training and support in teaching the material and in the use of R Studio. Several members of the teaching profession had tried to use R themselves with little success, mainly due to time constraints and accessing relevant learning materials. CPD events were again provided for teaching staff on basic R skills using the online resource produced by the Q-Step interns. This proved to be a successful way of training teachers to deliver the course with confidence and enthusiasm. Some of the positive feedback from the CPD training sessions included the following:

- *'I was an attendee at the Higher Statistics event last week. I would just like to thank you in terms of the delivery of unit information etc. Your knowledge and experience oozed and it was a very enjoyable day. Thank you.'*
- *'I particularly appreciated the way you gave real life experiences to back up statements and statistical theories.'*
- *'The time in the labs was very productive. I hadn't used Minitab since I was at University and it had certainly changed from the Windows Manager 3.1 version. Also changed days from looking up t tests and F tests values from books, that I remember.'*
- *'RStudio, I had never used before so to have hands on experience was very worthwhile. In terms of no cost this would be the one we would choose (for obvious reasons).'*
- *'I just wanted to thank you for the presentation on Thursday. I found it incredibly insightful and interesting. The demonstration on using R studio with be invaluable, as I was previously using Excel, which was unnecessarily cumbersome for some things.'*
- *'Going forward I will be "selling" the Stats Award to pupils as an invaluable set of skills for those planning on doing any course at university with a research element.'*
- *'Yesterday's course was very enjoyable. Thank you very much. I came away from yesterday's course very excited about doing statistics. I would like to consider doing more work on statistics at university level, what would be the best pathway for this? Any information would be appreciated.'*

3.5 Assessment

The Higher Statistics Award is assessed in a variety of ways and can include tasks and activities, for example:

- elements of practical assignments such as a project or investigation
- specific assessment tasks or activities
- discrete tests or question papers

These approaches can be used to provide a varied and integrated assessment experience and to make it more coherent and meaningful for learners. It can include learners being tasked to analyse a set of data using a variety of techniques, to answer questions from non-statistical users, and to present the analysis and conclusions in verbal or written form to address the aim of the investigation.

4. Discussion and Conclusion

The SQA Statistics Award introduces pupils in Scotland to a basic knowledge of statistics. More significantly, it allows them to develop skills in applying that knowledge to solve real life questions by interrogating data. The use of propriety statistical software packages like Minitab and R Studio allows candidates from an early age to start developing the computer literacy required to analyse and make sense of the vast amount of data which is available across all aspects of life from health-care to retail. The approaches to learning and teaching this unit promote skills in decision making, communication and critical appraisal of information which are key skills for learning, life and work. In addition, it provides a formal qualification for learners in these skills.

Presently the SQA are looking to incorporate the Statistics Award into a higher level qualification on Applications of Mathematics providing an important curricular pathway for young people. This would make the course more accessible to students across the country and is likely to be a useful mathematical qualification for students going on to further education in non-STEM subjects. This brings Scotland in line with other countries where statistics is taught at primary and secondary education level including New Zealand [13] and the United States [14].

Acknowledgements

The authors would like to sincerely thank the Scottish Qualification Authority for commissioning this work in an effort to address the gap in the school curriculum for education in data science and for funding CPD courses for Scottish teachers to deliver the award. We would also like to thank the teaching staff and pupils who provided invaluable feedback during the development phase of this work and their assistance in the choice of statistical software package. We are also grateful to Emmi de Vries and Kirsty English who were funded by Q-Step to develop the online teaching videos and resources for R Studio.

References

1. Drive launched to train 100,000 Scots in data skills, BBC News, September 2018. Available from https://www.bbc.co.uk/news/uk-scotland-scotland-business-45629533.
2. Royal Statistical Society, Education and skills for the data economy. Available from http://www.rss.org.uk/RSS/Influencing_Change/Data_manifesto/Education_and_skills_for_the_data_ economy/RSS/Influencing_Change/Data_democracy_sub/Education_and_skills_for_the_data_econo my.aspx?hkey=9cd80f64-6260-4502-97cf-6c1f47af2a72.
3. Transforming Scotland into a maths positive nation: final report of the Making Maths Count group. Available from https://www.gov.scot/publications/transforming-scotland-maths-positive-nation-final-report-making-maths-count/.
4. Leavy, A. and Hourigan, M. Crime scenes and mystery players! Using driving questions to support the development of statistical literacy. Teaching Statistics, 38, 1, p29-35.
5. Education and skills for the data economy. Available from http://www.rss.org.uk/RSS/Influencing_Change/Data_manifesto/Education_and_skills_for_the_data_ economy/RSS/Influencing_Change/Data_democracy_sub/Education_and_skills_for_the_data_econo my.aspx?hkey=9cd80f64-6260-4502-97cf-6c1f47af2a72#endnotes.
6. Independent Teacher Workload Review Group (2016) Eliminating unnecessary workload associated with data management (PDF). Available from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/511258/Eliminating-unnecessary-workload-associated-with-data-management.pdf.
7. Stokes, M. (2016) Making the most of data in schools (PDF), London: Nesta. Available from http://www.nesta.org.uk/sites/default/files/making-the-most-of-data-in-schools.pdf.
8. SQA Higher Statistics Award. Available from http://www.sqa.org.uk/sqa/73224.html.
9. Minitab Inc. Available from www.minitab.com.
10. SQA Advanced Higher Statistics. Available from https://www.sqa.org.uk/sqa/48506.html.
11. R Studio. Available from https://www.rstudio.com/
12. Stats Award. Available from https://statsaward.weebly.com/

13. Wild, C.J.; Pfannkuch, M.; Regan, M.; Horton, N.J. (2011). Towards more accessible conceptions of statistical inference. Journal of the Royal Statistical Society, Series A. 174 (2): 247– 295.
14. Lindsay, Bruce G.; Kettenring, Jon; Siegmund, David O. (August 2004). A Report on the Future of Statistics. Statistical Science. 19 (3): 387–407 (page 403).

Differences in Perceptions towards STEM of Male and Female Students in Australia

Elizabeth Stojanovski[1], Inge Koch[2]
[1] University of Newcastle NSW Australia
[2] The University of Melbourne VIC Australia

Abstract
There appear to be differences in mathematics literacy performance between genders based on national studies, including the 2012 International Programme for International Student Assessment (PISA), with males outperforming females, and females with higher levels of anxiety towards mathematics than males. In Australia there is a shortage of employees in science, technology, engineering and mathematics (STEM) related careers. In addition, there is a strong underrepresentation of females in these areas, similar to the trends found among females compared with males studying mathematics in later years of secondary school and in STEM tertiary level studies.

This paper investigates data from secondary school students participating in the 2012 and 2015 PISA to assess attitudes and perceptions of students towards mathematics. Differences in attitudes towards mathematics between genders are investigated. Based on these findings, it is proposed that different learning strategies are needed. Such strategies would require the implementation of different teaching practices in an attempt to reduce student anxieties towards mathematics. An increased emphasis on presenting to students real world applications of mathematics and statistics is also suggested so that students are better aware of the high demand for such skills in the workforce. This changed focus could potentially impact on the number of students and, in particular, female students choosing to study mathematics and statistics in tertiary education to help ensure greater equity and representation of females in STEM related areas. There is also a need to transfer these skills and knowledge to teachers in the secondary education setting so that students in the secondary school setting are better informed. This could enable students to be better equipped to consider taking relevant courses during their final years of secondary study to hence enable these students to be better prepared for tertiary study in STEM areas.

Strategies are suggested to help to improve the shortage of STEM qualified employees. The views of what it means to be a capable mathematician could be modified to encourage more females to study mathematics and statistics. This could have broader implications in terms of government national policies and incentives to entice further education in and participation in STEM areas while also addressing the shortage of females in these areas.

Keywords
Mathematics education; Gender; STEM; Careers; Equity

1. Introduction

It has been reported in Australia that there is a shortage of employees in science, technology, engineering and mathematics (STEM) related careers [1]. There is also a strong under-representation of females in these areas, similar to the trends found among females compared with males studying mathematics in later years of secondary school and in STEM tertiary level studies.

The 2015 Programme for International Student Assessment (PISA) study results released by the Australian Council for Educational Research show that Australian year nine student mathematics literacy results declined by 30 points since 2003, equivalent to one year of schooling, with 10 points of the decline occurring in the last three years [2]. It has been claimed that Australian students' ability to apply their mathematical and scientific knowledge to real life situations is falling both absolutely and relatively compared to other countries [3]. Differences in mathematics literacy performance between genders have also been observed in Australian national studies, including the 2012 PISA study, with males found to outperform females [4]. Some other studies, however, have shown school grades of more than 1.6 million students from around the world, with the majority from North America, in STEM subjects to be similar between male and female students, with the top 10% of a STEM classroom containing equal numbers of females and males [5]. The small gender gap noted in some studies in ability does not seem to explain the large gender gap in terms of undergraduate students choosing to study mathematical STEM subjects and working in STEM careers [6].

A proposed contributing factor to this trend, however, is gender differences in self perceptions with females underrating their abilities on tasks and careers considered male dominated [7]. It has been shown, for example, that among students performing at the 90th percentile in mathematics ability in late secondary school, boys still rate themselves higher than do their female peers with the claim that girls who are more confident in their mathematics skills are more likely to pursue mathematics based degrees [7]. It was found in one study that year 12 females with the highest levels of confidence in their mathematics ability were more than three times more likely to major in math intensive STEM areas than females with the lowest levels of confidence [7].

Another contributing factor postulated to contribute to this difference is that girls are more anxious about mathematics [8]. Students who have high levels of mathematics anxiety tend to underperform in mathematics tasks compared to students with lower levels of mathematics anxiety [9]. It has also been claimed that those who experience mathematics anxiety tend to avoid mathematics, mathematics courses and career paths that require mathematics skills [10].

This paper investigates data from secondary school students participating in the 2012 and 2015 PISA studies to assess gender differences in attitudes, learning methods and anxieties of students towards mathematics.

2. Methodology

Data Sources

Data will be used from the Programme for International Student Assessment (PISA) study which is intended to assess educational systems by measuring 15-year-old school students' scholastic performance on mathematics, science, and reading and is held every three years. The PISA study is a comparative study of student achievement directed by the Organisation for Economic Co-operation and Development (OECD), among students who are nearing the end of their compulsory schooling in most participating educational systems to determine how prepared they are to use the knowledge and skills in particular areas to meet real-life opportunities and challenges [4]. PISA assesses a random sample of 15-year-old students, drawn from a nationally representative sample of schools. In Australia and is managed by the Australian Council for Educational Research (ACER) [4].

2012 PISA Dataset: For the present study, data from 14 481 secondary school students participating in PISA 2012 study in Australia were analysed in terms of their mathematics anxiety, which formed a focus of the 2012 study. The study also measured students' mathematical literacy in terms of several mathematical techniques (including the mathematical formulation of scenarios; the use of mathematical concepts and methods in mathematics; and the interpretation and application of mathematical solutions) and topics (including relationships; aspects of space and shape; quantification; data and uncertainty) [4]. Mathematics performance was measured using a total score representing the number of administered questions, covering all of these mathematical techniques and topics, which were answered correctly. The total scores were used as the main outcome variable for the present study. The questionnaire developed by PISA also contains information on mathematics anxiety with students asked to report their level of agreement with two items relating to mathematics anxiety: 'worry that mathematics classes will be difficult', and 'get very tense when they have to do mathematics homework'. The possible responses options for each question were presented as four categories: strongly disagree, disagree, agree and strongly agree. Student responses about their feelings of stress associated with anticipating mathematical tasks and while attempting to solve mathematics problems were used to identify students' specific level of anxiety towards mathematics

2015 PISA Dataset: The 2015 study was the sixth cycle of PISA since it was first conducted in 2000. The present study presents results for Australia from a total of 14 530 student participants with 7163 females and 7367 males. Australia took a larger sample than required in order to oversample smaller jurisdictions and Indigenous students to ensure reliable estimates. The main mode of assessment changed from a paper-based delivery to a computer-based delivery with students completing a series of questions with stimulus material comprising a short written passage or text accompanying a table, chart, graph, photograph or diagram. A range of item-response formats including multiple-choice and open ended responses were used to cover the full range of cognitive abilities and knowledge identified in the Assessment Framework.

In addition to the standardised student questionnaire that students were assigned, an additional student questionnaire offered was the educational career questionnaire which sought information from students about their learning habits and preparation for their future career with 6863 students completing the additional career questionnaire, representing 47% of the original sample. One question asked about additional instruction in school subjects attended by students in the school year. This instruction could take place at school or somewhere else, but is not part of students' mandatory school schedule and students were asked to consider all regularly attended, institutionalised, organised additional learning activities in which they received some kind of instruction, guidance, or support and to approximate how many hours per week they attended the additional instruction in mathematics.

Statistical Analyses

Variables were used on two different levels in the proposed analyses, with student mathematics anxiety measures, gender and preferred learning methods measured on the individual student level, while school was used at class level. The effect of gender on anxiety and preferred learning methods will be assessed descriptively using percentages and chi-squared tests and will be validated using multilevel binary logistic regression models, also known as generalised linear mixed models, to take into account the hierarchical structure of data with students nested within school groups, and simultaneously model data from the different levels. Multilevel regression models will be used to assess differences in time spent in additional instruction in mathematics between genders.

3. Results

Mathematics Anxiety by Gender

A considerable proportion of students reported feeling stressed when dealing with mathematics on the 2012 PISA study. Across students in Australia, 60% of students reported that they often worry that it will be difficult for them in mathematics classes. There was a wide variation between genders on this item with 68% of females feeling worried (agree or strongly agree with the statement) that it would be difficult for them in mathematics classes compared with 53% of males ($p<0.01$) (Figure 1). Of students completing the questionnaire, 38% reported that they get very tense when they have to do mathematics homework, with wide differences in levels of mathematics anxiety by gender with girls reporting stronger feelings of mathematics anxiety than boys (41% and 35% respectively for females and males, $p<0.01$).

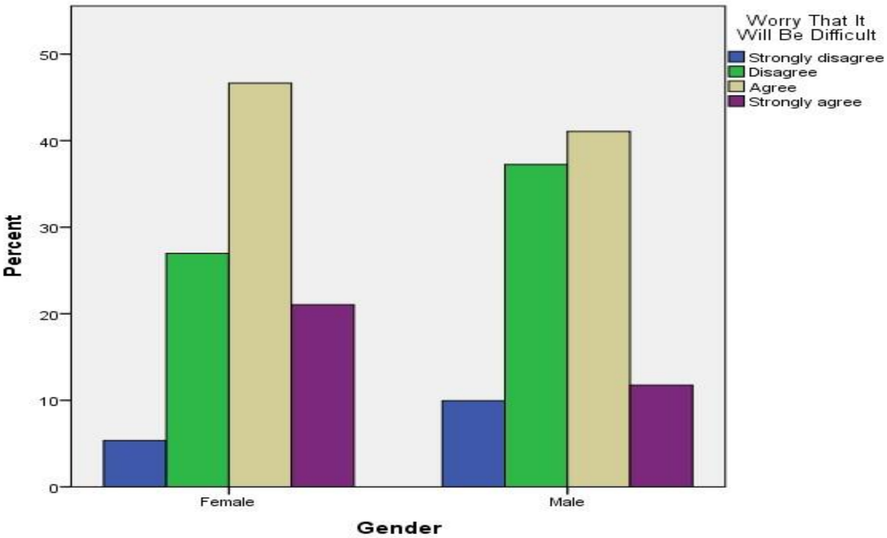


Figure 1: Bar Chart displaying the percentage of students who worry that Mathematics classes will be difficult by Gender.

Study Habits and Perceptions by Gender

Males appeared to spend more hours per week (Mean: 3.83, 95% CI: 3.69-3.97; Median: 3, IQR 4) attending additional instructions in mathematics, compared to females (Mean: 3.1, 95% CI: 2.97-3.22; Median: 2, IQR 4) ($p<0.01$). Students were also asked to approximate how many hours per week they spend learning in addition to the learning that takes place in the classroom in mathematics. Females appeared to spend less time, on average (Mean: 4.71, 95% CI: 4.56-4.87; Median: 4, IQR: 3), compared with males (Mean: 5.45, 95% CI: 5.26-5.65; Median: 4, IQR: 4) learning in addition to that which occurred in the classroom ($p<0.01$). These results indicate potentially higher degrees of motivation among males to study mathematics beyond what is covered in the classroom.

Interestingly, among those who did spend additional time in mathematical instruction, substantially more males than females indicated that the additional mathematics instruction covered additional content that was not covered in regular school courses (49% of males compared with 40% of females, $p<0.01$). These results also suggest that males appear more motivated and perhaps more able to do extension work beyond what is covered in the regular classroom.

Table 1: Reasons for attending additional mathematics instruction by Gender

Reason	**Gender**	**%**	***p*-value**
Wanting to learn	Males Females	52 47	<0.01
Attracted by tutoring ads	Males Females	18 10	<0.01
Parents recommended tutoring	Males Females	68 66	0.04
It was the teacher's recommendation	Males Females	27 21	<0.01
Many friends are doing it	Males Females	22 15	<0.01
It was gratifying to study	Males Females	22 16	<0.01
It looks good on a resume	Males Females	23 15	<0.01
Necessary for a job	Males Females	27 20	<0.01
Want to prepare for exams	Males Females	45 49	<0.01
Want to improve grades	Males Females	44 49	<0.01
Need to improve grades	Males Females	36 34	0.08

Males were significantly more likely than females to attend additional mathematics instruction for the following reasons (Table 1): wanting to learn, attracted by tutoring ads, Parents recommended tutoring, teacher's recommendation, many friends are doing it, it was gratifying to study, it looks good on a resume and necessary for a job, suggesting that even at age 15, males appear to potentially be thinking and preparing for STEM related courses of for tertiary level study or career more than females. Females, on the other hand, were significantly more likely to attend additional mathematical instruction to: prepare for exams and to improve grades.

Furthermore, there appeared to be differences in the type of study and learning preferences by gender (Table 2). Significantly more males than females participated in the following mathematics instruction types: internet tutoring with a person (such as skype), internet or computer tutoring with a program or app, video recorded instruction by a person, video recorded instruction by a person, and large study groups, indicating potentially greater self-motivation among males compared with females (Table 2). Contrastingly, slightly more females than males indicated their learning occurred was via small group study or practice groups comprising between 2 and 7 students suggesting a more personalised learning style.

Table 2. Type of Study and Learning Preferences by Gender

Type	**Gender**	**%**	***p*-value**
Internet tutoring with a person (such as skype)	Males Females	11 6	<0.01
Internet tutoring with a program	Males Females	15 10	<0.01
Video recorded instruction by a person	Males Females	11 6	<0.01
Small groups study (2-7 students)	Males Females	27 25	0.07
Large group study (8 or more people)	Males Females	19 16	<0.01

4. Discussion and Conclusion

These results suggest that males appear more self-motivated, compared with females, in terms of participation in additional mathematics study beyond the classroom on topics not necessarily covered in the classroom and appear less anxious than females in mathematics. Males also demonstrated greater self-motivation by participating in more online learning methods than females and appeared more comfortable in large study groups, compared with females, possibly due to having more friends also participating in these. Females, on the other hand, appeared to prefer, although not significantly more than males, small study groups potentially preferring the one on one more personal study technique. Also apparent was that male students were more likely to participate in additional instruction as a result of recommendations by teachers and parents or to better prepare for jobs while females were more likely to select better preparation for exams and to improve grades, without having a longer term goal in terms of career prospects.

Improving the views of what is meant by a capable mathematician or statistician could encourage more females to study mathematics and statistics subjects and participate more in STEM based careers with a mathematics focus. Gender equity should try to make female participation in mathematics a priority by changing the views of a male dominated area. Improved support from mentors such as teachers and parents may help counteract what appears to still be a male dominated field in order to encourage girls to feel confident and less anxious to succeed in mathematics areas. This could also have implications for national policies to offer incentives to help promote female participation in STEM male dominated areas via secondary school teachers potentially redesigning courses and assessments to help promote female student participation and confidence as well as initiatives aimed at parents as they also influence student career paths, while also addressing the shortage of females in these areas.

References

1. Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., and Byers A H. (2011). Big data: The next frontier for innovation, competition, and productivity. Retrieved from: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.
2. Karp P. (2016). Pisa results: Australian students' science, maths and reading in long-term decline, The Guardian. Retrieved from: https://www.theguardian.com/australia-news/2016/dec/06/pisa-resultsaustralian-students-science-maths-and-reading-in-long-term-decline.
3. Thomson S., De Bortoli L., Underwood, C. (2017). Programme for International Student Assessment (PISA). PISA 2015: Reporting Australia’s results by Australian Council for Educational Research Ltd www.acer.org/ozpisa/reports/ © Australian Council for Educational Research Ltd.
4. OECD (2013). PISA 2012 Results: What Students Know and Can Do (Volume I): Student Performance in Mathematics, Reading and Science OECD Publishing.
5. O’Dea R.E., Lagisz M., Jennions M.D., Nakagawa S. (2018). Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. Nature Communications. 9, 3777.
6. Wang M. T. and Degol J. L. Gender gap in science, technology, engineering, and mathematics (stem): current knowledge, implications

for practice, policy, and future directions. Educational Psychological Review 29, 119–140 (2017).
7. Perez-Felkner L., Nix S. and Thomas K. (2017). Gendered Pathways: How Mathematics Ability Beliefs Shape Secondary and Postsecondary Course and Degree Field Choices. Frontiers in Psychology, 8(386). doi:10.3389/fpsyg.2017.00386
8. Kyttala M. and Bjorn P. (2010). Prior mathematics achievement, cognitive appraisals and anxiety as predictors of Finnish students' later mathematics performance and career orientation. Educational Psychology 30, 431-448.
9. Ma X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. Journal for Research in Mathematics Education 30, 5, 520-540.
10. Ashcraft M.H. and Ridley K.S. (2005). Math anxiety and its cognitive consequences, in J.I.D. Campbell (ed.), Handbook of Mathematical Cognition, Psychology Press, New York, pp. 315-327.

The balance between governance support needed and influence avoided; the case of population censuses in developing countries

Jean-Michel Durr
National Institute of Statistics and Economic Studies (Insee)

Disclaimer
The views and opinions expressed in this article are those of the author and do not necessarily represent those of Insee.

Abstract
Population and housing censuses are large operations mobilizing the whole country. The National Statistical Office, usually in charge of the preparation and implementation of the census, does not have the permanent resources to conduct it on its own and needs support from the national government as well as from local authorities. There are three areas where the government plays a role in a census. It provides the legal framework to conduct the census; funding for the census; and logistical support. Local authorities provide also logistical support and, in some countries, a 'democratic control' of the fairness of the enumeration operations. However, the involvement of national and local authorities in the census operations is not without pitfalls. The independence of the census may be jeopardized, and the census perceived by the population as an administrative operation rather than a statistical one. This may entail lack of trust from the population on the use of the data and the confidentiality guaranteed to the information provided. Based on examples in the Western Balkans, Asia and Africa, the paper discusses the optimal balance between governance support and influence avoided in the population and housing census operations.

Keywords
Population and housing census; quality; governance

1. Introduction

According to the Fundamental Principles of Official statistics [1], the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data. However, population and housing censuses are large operations that usually cannot be conducted solely by the National Statistical Office without support of the Government. According the UN, there are three areas where the Government plays a role in a census [2].

Firstly, the Government provides the legal framework to conduct the census. A comprehensive framework, entrusting the census agency to carry out census activities, providing for the role of government departments in the operation, and including provisions related to the obligations of the participants as well as confidentiality safeguard measures is needed to conduct the census in a proper way.

Secondly. Government provides funding for the census, at least partially. Agreement by Government on the level of funding for the census is needed early in the cycle, from preparation to data collection and processing, so that other aspects of census planning can proceed.

Thirdly, the Government provides logistical support to the census. For example, in some countries, teachers are mobilized as enumerators, with the support of the Ministry of Education. Ministry of Justice provides access to the prisons for the enumeration of prisoners. Other government agencies sometimes supply specialist services such as form printing, mapping, transport services or media liaison. At local level, local authorities are called upon to provide training rooms and equipment, and meeting space for field staff.

In this context, the population and housing census can be considered as an administrative operation, under the control of the Government, rather than as strictly statistical operation. The perception of a census as a largely administrative operation could lead to reluctance of the population to participate in a governmental operation, people not trusting the neutrality of the operation and the confidentiality of their responses.

2. Government or political interference in census operations

However, it may happen that the government goes beyond its supporting role as defined above and attempts to interfere in the technical organisation and conduct of the census. Political leaders may also tend to interfere with census operations in relation with their political agenda.

Interference can be defined as actions to prevent a process or activity from continuing or being carried out properly. Objectives may be to influence the results of the census, for example in terms of size of the population at national or sub-national level, or in terms of the relative weight of population groups (religious or ethnic groups, for example) or seize opportunity of the census to highlight its achievements.

Government or political interference, as defined above, can take various forms: introduction or removal of specific questions; exclusion of specific population groups or inclusion of non-resident groups; alteration of the results during the processing phase; delays or non-dissemination of the results. Several steps of census operations can thus be affected. Based on observations or in documentation provided by international monitoring operations, we will give examples of such cases in the various phases of the census operation.

Questionnaire design

Whereas it is normal for the Government to express its needs for topics to be included in the questionnaire for considerations of policy making, it can happen that the Government or political leaders try to interfere with the census operation by influencing the content of the questionnaire. In some countries the census questionnaire is submitted for approval to the Government, leading to possible interference as in Canada when the Government decided to abolish the obligation to answer to the long form of the census 2011.

Despite the risk to jeopardize the census due to the sensitivity of the ethnicity topic, Western Balkans countries governments pushed to include such topic, as well as religion and mother tongue, in the census questionnaire. The most illustrative example is North Macedonia, where the distribution of ethnic groups has a direct impact on the civil rights of the population groups in application of the Orhid agreement. Despite warnings that asking for ethnic affiliation in the 2011 census questionnaire would lead to inevitable tensions among communities and frauds, the Government maintained its request to include these sensitive questions in the census. The disorders and ethnic tensions that these sensitive questions triggered during the fieldwork forced to stop the operation after ten days into the enumeration period. Nevertheless, it seems that the next 2020 census will still include these questions. This example reveals the naive belief that a statistical operation can collect any information, regardless its sensitivity, which is of course false. The paradigm governing statistical activities is that respondents have no interest in lying or in tampering with their answers, but if their response may have a direct consequence on their civil rights, it is likely to generate frauds.

In Bosnia and Herzegovina, the formulation of a question on the ethnic affiliation raised issues. International monitoring observers advised to stick to an open-ended question, as recommended internationally, to allow individuals to answer such questions freely, while political pressure pushed to ask the question using predefined boxes proposing the three ethnicities “constituents” of the population according to the constitution, and an option “other, specify”. A compromise was reached, but this was not in full compliance with international recommendations.

More recently, the request of the US Government to include a question on citizenship in the census 2020 raised concern about the risk that undocumented migrants might decline to respond to the census. By law, the Census Bureau, part of the Department of Commerce, is not allowed to share individuals’ data with law enforcement agencies, but some immigrants may nevertheless have that fear. Because census data is used to apportion seats in the US House of Representatives, dole out federal funding and serve as the basis for state-level political districts, an undercount of immigrants could affect a state or city's political representation for a decade. The case came to the Supreme Court who rejected the commerce department's stated justification for including a question on citizenship.

Field work

Instructions given by the Government during fieldwork sometimes interfere with the conduct of the census. In Myanmar, the last 2014 census was marred with serious irregularities regarding the implementation of the international methodological and human rights standards. International recommendations clearly require the option for self-declaration of ethnicity and the indication of multiple ethnic affiliations (as opposed to using pre-established classifications). The Government of Myanmar listed 135 predefined official ethnic categories, comprising eight major groups and sub¬groups. However, this list did not enjoy consensus among ethnic leaders. Early on in the preparations, UNFPA advised the Government to ensure that self-identification would be respected for all. Nevertheless, a few days prior to start of enumeration, the Government reneged on its commitment to allow for the self-identification of all ethnic groups, as it refused to enumerate those who would declare Rohingya as their ethnicity. As a consequence, the census did not cover certain areas of Kachin and Kayin States as well as in Northern Rakhine, where communities were not allowed to self-identify as "Rohingya". Analysis of census information by the Department of Population, in charge of conducting the census, revealed and estimated total of over 1,200,000 not enumerated in parts of the Rakhine State, Kachin State and Kayin State. This represents 2.34 percent of the population [3].

Less tragic is the tentative to inflate the population counts of a municipality or region for financial or political reason. The census in many countries is the basis on which apportion the sharing of funds or seats at the Parliament, or even representativeness in the national Government. Thus, fraud attempts are observed in many countries, developing as well as developed, in particular when local authorities are involved in census operations. Census Commissions, established in several countries, are often problematic in that regard. Depending on their composition, their role can include: logistics support but in some cases also "democratic" control of the census operations. Composed of local citizen but also political or ethnic groups representatives, they sometimes exceed their mandate, giving direct instructions to census field workers, in contradiction with the methodology, for example to enumerate persons who are not resident.

The census communication campaign can also be used by the Government to promote its achievements or policy, with the risk to reinforce the perception of the population that the census is an administrative operation conducted by the Government.

Processing

Governmental interference with the processing of census data is more complex to witness. However, instructions given to the census agency regarding the process of data editing can have consequences on the census results, including population counts. In Bosnia-Herzegovina, the three statistical agencies (Statistical office of the Federation of Bosnia-Herzegovina, Statistical office of Republika Srpska and National Agency for Statistics) did not agree on the way to determine the resident status of enumerated persons, based on their responses to several probing questions in the 2013 census form. The situation with the three statistical agencies ultimately escalated into the political and public debate, receiving broad attention in the country. After more than one year of discussion, the Director of the Agency for Statistics took the decision to implement the recommendations of the International Monitoring Operation [4], which was still contested by one of the entities statistical office, who decided to perform its own data processing and publish its own figures, which at the end revealed to be close to the official ones.

Dissemination

Political interference can also occur during the dissemination phase, due to the importance of census figures in the political and social context of a country. As census population counts are often used to apportion seats in the Parliament, or share public funds, they are of high importance for political leaders. In addition, ethnic or religion distribution may have consequences on the social balance of influence in a country.

As evidence of lack of independence of the census agency in the release of census results, some countries have the practice to seek formal "approval" by the President of the Republic on the census population counts, and examples exist of non-approval leading either to non-publication of the results or even possible alteration of the counts. Delay in the release of results is also a way of depreciating census results.

The results of the census 2005 of Cameroon were released only in 2010 and were disputed. Census results, especially the distribution between the Muslim North and the Christian South, have often sparked controversy in Nigeria, and the results of the 2006 census were no exception [5]. The provisional results of the 2006 census were released by the President in January 2007 and immediately criticized, both on the level of the total population, just above 140 million, and on the distribution of state population [6].

Kenya is an interesting case of politico-judicial interference in the census process. During tabulation of the 2009 census, it became clear to Kenyan population specialists that the population figures were higher than expected in eight counties in the North Eastern Province.[1] Due to failure of the post enumeration survey in this region, the problem was unfortunately not detected earlier. During the official release of the census results in 2010, the Minister for Planning, National Development and Vision 2030 nullified the census results in these eight counties and announced a recount of the area to obtain more accurate population data. The recount was barred by the High Court after members of parliament from the area filed a case. The judge in the case decided that the actual enumerated data were the official figures and should form the basis for further analysis. This had serious consequences for the subsequent use of census data.

According to Goble [7], Turkmenistan's Government did not release the results of the 2012 Census that seems to have revealed that the actual population was far less than official claims and the ethnic composition too differed from that expected by the Government.

Data on religion of the census 2011 of India were released only in August 2015. Observers, both within and outside India, alleged that the 2011 Census data on religion were not released due to electoral considerations [8].

Confidentiality

Strict confidentiality safeguards are needed to protect census data from political interference. It has happened that government officials, or political parties request or exert pressures on the Statistical Office to access to individual census forms or records. In Montenegro, the Director General of the Statistical Office faced a lawsuit because it refused to grant access to census forms to a political party. Fortunately, the Court ruled in favour of the Director of the Statistical Office. However, it shows the lack of respect of statistical confidentiality that some politicians display, but also the difficulty to educate people that a population census is a strictly statistical operation, and not an administrative one.

3. How to maintain the census away from political or undue government interference?

The UN Principles and Recommendations for Population and Housing Censuses [9] highlight the importance that national statistical/census agencies

[1] Lagdera, Wajir East, Mandera Central, Mandera East, Mandera West, Turkana Central, Turkana North and Turkana South.

maintain professionalism and demonstrate neutrality and objectivity in the presentation and interpretation of the results and are free from real or perceived political interference so that the objectivity and impartiality of the statistics is assured, because this in turn will build trust in and acceptance of the results.

The question is thus how avoid political interference in census operations. The first response is to have appropriate legislation ensuring the independence of official statistics. This is even more needed as in many countries, the National Statistical Office (NSO) or the Census Agency is part of the public administration, organised/subordinated to under the responsibility of a line minister. European Member States have to comply with the European Statistics Code of Practice, of which first principle is professional independence, meaning that the independence of the National Statistical Institutes from political and other external interference in the production, production and dissemination of statistics is enshrined in law and guaranteed for other statistical authorities[2]. Statistics acts in various countries, as South Africa, include provision related to the sole responsibility of the Chief Statistician to determine, and exercise final responsibility regarding the data collection, methodology, and dissemination of statistics.

This entails that the National Statistical Office or the Census Agency should be entrusted with the full responsibility of conducting the population and housing census, including designing, conducting the operation, processing and disseminating the results. National Census Commissions, when existing, should restrict their role to support to the census, but not to decide on the content of the census questionnaire and the methodology. In North Macedonia, the Statistical Office was by law deprived of full responsibility for the census 2011. The law provided that "the Census is organized and conducted by the State Census Commission in co-operation with the State Statistical office". The Commission felt entitled to challenge the methodology proposed by the State Statistical Office, which was in compliance with international recommendations, creating a confusing situation on the ground, leading to the cancellation of census operations after ten days.

This of course does not mean that the Statistical Office shall decide on its own on the content of the questionnaire and not organize stakeholders' consultations. In that regard, Statistical Councils composed of representatives of users, from governmental departments, the public sector, civil society organizations and the private sector, are useful institutions to ensure that the

[2] European statistics Code of Practice - revised edition 2017. Available from: https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118-f729c75878c7

topics included in the census questionnaire, as well as other statistical operations, are fully relevant.

At local level, the need for census commissions should be carefully evaluated. Are they really needed? Logistical support can be carried out by one person responsible at local authority level to provide training facilities, storage space, and support fieldworkers in case of refusals to respond to the census. The "political" control of the enumeration, sometimes entrusted to local census commissions, is highly questionable, as it insinuates that the NSO may not be reliable to conduct the census in a neutral way, thus raising suspicion about the operation.

International organisations have also a role to play. They support censuses in many developing countries, organizing resource mobilization and providing technical assistance. This gives them the right and ability to publicly denounce misconducts and even withdraw support, but they rarely use it. Stronger donor involvement in ethics and compliance with international standards in the conduct of census operations would also help mitigate political interference in the censuses they support.

4. Conclusion

While promoting and supporting the population and housing census, the Government must adopt a neutral attitude, refraining from taking the lead over the operation, with the risk that the census would be seen by the population as a governmental operation. The Government should communicate on the importance of the census for the country but emphasizing the role of the NSO in designing and conducting the operation. It is also very important to clearly separate the census from other operations, such as using the census to build an administrative register, unless this is explicitly provided by the census law, or asking the enumerators to collect information for additional survey, for example on the opinion of the public on a governmental project. The publication of the census results should be the sole responsibility of the NSO or Census agency, without any "approval" from the Government, even where the population counts are authenticated in the official gazette.

Finally, the role of the media is key in a democracy. They must be informed early enough of the privacy and neutrality purpose and safeguards of the census, to ensure that enumeration and subsequent steps are conducted in accordance with these principles. Regularly communicating with the media is the best way for the NSO to ensure transparency throughout the census operation.

References

1. Fundamental Principles of Official Statistics. Resolution adopted by the United Nations Economic and Social Council on 24 July 2013. E/RES/2013/21. New York. 2013.
2. United Nations Handbook on the Management of Population and Housing Censuses, rev. 2. Final draft, unedited, not typeset. New York. Available from: https://unstats.un.org/unsd/publication/seriesF/Series F83Rev2en.pdf
3. The 2014 Myanmar Population and Housing Census. The Union Report. Census report volume 2. Department of Population, Ministry of Immigration and Population. May 2015.
4. Hoh, A-L (2017). 'When counting counts' - Europeanisation of census taking in Croatia, Bosnia and Herzegovina and the Former Yugoslav Republic of Macedonia. Journal of Contemporary European research. Volume 13, Issue 1 (2007).
5. Yin, S (2007). Objections Surface over Nigerian Census Results. PRB. Accessed from: https://www.prb.org/objectionsovernigeriancensus/
6. Akinyoade. A et alii (2017). Census-taking in Nigeria: the good, the technical, and the politics of numbers. African Population Studies. Vol. 31, No. l, (Supp.).
7. Goble, Paul (2015). "Unpublished Census Provides Rare and Unvarnished Look at Turkmenistan." Eurasia Daily Monitor, Volume: 12 Issue: 26.
8. The Economist (2013). "'India's Muslims Growing, and neglected." March 02.
9. United Nations (2015). Principles and recommendations for population and housing censuses the 2020 round, rev. 3. Series M. No.67/Rev. 3. New York.

Governance models and their impact on quality in the statistical system

Delina Ibrahimaj
Director General of Institute of Statistics Albania, Tirana, Albania

Abstract

Pursuant to national legal frame, the public institutions in Albania are divided in two categories: governmental institutions and independent institutions. The Institute of Statistics in Albania (INSTAT) has been part of the first category until May 2018, under Prime Minister. Aiming to increase the perception of the public and users on the independence of statistics produced, implementing at the same time the recommendations of Eurostat for increasing the professional independence level, INSTAT proposed to the government the new institutional positioning. In 13th May 2018 a new law on official statistics entered into force, defining INSTAT as a public independent institution charged to report directly to Albanian Assembly.

The new positioning of the institution in a higher hierarchy level improved the role of INSTAT as coordinator of National Statistical System (NSS) and the mandate of INSTAT for accessing administrative data by prevailing over any provision provided in other laws regulating administrative data. In addition, it increased the autonomy in managing of human and financial resources. INSTAT has the full competence to observe the standards, concepts, definitions, classifications and statistical methodologies used by the other statistical agencies, in order to ensure the quality in official statistics. The overall quality of statistical products or services will be evaluated taking into account the quality aspect such as: timeliness, completeness, comparability and trust. User Satisfaction Survey will be used to measure the degree to which INSTAT meets user's needs. The Staff Satisfaction Survey will be used to identify gaps and needs to have more engaged staff in order to increase the institution culture.

This paper will be viewing the following dimensions of the governance: political system, legislation, impartiality, transparent administrative procedures, feedback from users, quality, effectiveness and efficiency.

The paper will provide also detailed information on the impact of the new positioning of INSTAT as an independent institution towards the confidence of users and general public in official statistics.

Keywords

Independence; Reliability; Trust; National Statistical System

1. Introduction

The notion of 'governance' is broad and complex, applied in a variety of context, and encompassing both technical and substantive aspects. [1]Even though there is a general agreement on the importance for governance models, there is no clear definition on what it is. Jeston & Nelis (2008) summarize governance as an instrument to guarantee that strategies, processes and their design are performing well, and to ensure alignment among all three.

This paper will be viewing several dimensions of governance. External and internal factors impacting the governance model will be described and their impact on the quality in the National Statistical System (NSS) will be evaluated.

For the purpose of the study the external factors impacting the governance model which will be analysed are: the mandate and independence of the statistical office and NSS, national legislation and EU accession requirements impact. While regarding the internal factors, the most important ones taken into account are: INSTAT steering body, budget management, human resources management, efficiency and effectiveness in statistical production.

Within the Albanian legal framework the National Statistical System is composed of the official statistics producers, operating within the Republic of Albania in collecting, processing and publishing official statistics according to the Official Statistics Program (OSP)[2] . INSTAT is positioned at the centre of the Albanian national statistical system and the law on Official Statistics guarantees its mandate on the production of statistics and coordination of the statistical system. In 2018 a new law on official statistics was presented and approved by the Albanian Parliament. This law has increased the hierarchy level of the institution among public administration by making INSTAT an independent institution, reporting directly to the Assembly. In addition, the law empowered INSTAT by increasing its role of coordination of the national statistical system and by providing a clear mechanism of coordination and enforcing the role of the General Director of the NSI in the coordination of the NSS. This change directly affected the external governance system of the Institute of statistics. In terms of internal Governance the new law in official statistics provided for the creation of an internal steering body composed by directors of the institution with specific obligations in terms of decision making. Last but not least the new law in official statistics provided for restructuring of the Statistical Council mainly composed by personalities in the field of statistic from statistical agencies, civil society, media and academic world. This body is the highest body in hierarchy to supervise and support NSS

[1] IN-DEPTH REVIEW OF GOVERNANCE STATISTICS IN THE UNECE/OECD REGION, CES ECONOMIC COMMISSION FOR EUROPE 23 September 2016

[2] Law No.17/2018 on Official Statistics, http://instat.gov.al/media/3972/law-no17-2018-on-official-statistics.pdf

in the implementation of the OSP. The composition of Statistical Council will help INSTAT to ensure that quality and professional standards meet user's needs.

The Albanian NSS is composed by three Statistical Agencies: INSTAT, Bank of Albania and Ministry of Finance. As the coordinator of the NSS INSTAT, must develop methodologies, concepts, definitions, standards and classifications that need to be adopted by the other statistical agencies, wherever possible, while producing official statistics. A five year OSP is approved by the Assembly, and includes the statistical production of the entire NSS. Common reports and dissemination calendar, for the NSS, are coordinated by INSTAT. INSTAT in the framework of fulfilling obligations deriving from OSP has been working to increase the usage of administrative data for statistical production. In this regards memorandums of understanding (MoUs) with administrative data owners as well as research community have been signed in the last years.

INSTAT is committed to quality and this is not related only to the statistics that are produced within the institute, but also to the guidance and monitoring towards the quality assurance of production of official statistics by the National Statistical System (NSS). High quality in official statistics is achieved through a well-functioning and coordinated statistical system, which is fundamental in the efforts toward fulfilling the requirements of the European Code of Practice (CoP).The new law has increased INSTAT power in terms of coordination of the statistical programming and reporting, monitoring quality, methodologies, data transmission and communication on the activities of the NSS. It is INSTAT General Director's duty to guide the coordination of NSS, decide on methodologies, standards and procedures of official statistics produced and request for statistical audits on statistical production processes within the NSS, when appropriate.

In the work towards building trust in official statistics and a quality culture for the statistical processes and products, there is a need to have a mutual mission and a cooperative working environment in the scope of NSS. INSTAT has developed a Training Strategy including trainings of Statistical Agencies in methodologies, concepts, definitions, standards and classifications which will influence the quality improvement, by decreasing rates of noncompliance with the European CoP and improvements in customer satisfaction.

The Steering Body is the governing body of INSTAT, composed by the mid and high level management of INSTAT, chaired by the Director General. Its responsibility is to propose to the Director General, guidelines, methodologies and standards to be implemented by the NSS in order to ensure quality for collecting, processing and publishing official statistics. The new role of the Steering Body will facilitate the decision-making mechanism to strength the professional and institutional independence.

INSTAT and NSS are part of the transformation processes towards EU integration (Stabilization Association Agreement (SAA)). In the process of fulfilment of the requirement of SAA, INSTAT and statistical agencies should produce statistics in a transparent, impartial and timely manner, helping users to judge the performance of the transformation of the country.

2. Methodology

The paper will study the effect of the new law through internal and external factors affecting the governance of institution by analysing quality aspects.

External and internal factors impacting the governance model has been taken into consideration. In regards to the external factors, the impact of the new law on Official Statistics, approved in 2018, was measured to evaluate the mandate and independence of the statistical office and NSS. User Satisfaction Survey (USS) was used by INSTAT to understand users perception regarding quality in official statistics, as explicitly addressed in the European Code of Practice(CoP) "...to show that European and national statistical authorities are impartial and that the statistics they produce and disseminate are trustworthy, objective and reliable" (European Commission, 2005)[3] . The following dimensions have been evaluated: Satisfaction with the employees; Satisfaction with data; Satisfaction with the website; Satisfaction with the quality of data and services; Trust – INSTAT has reliable data; Overall Index. For this paper the following sub-dimensions have been studied: Statistics are distributed equally to users, Accuracy and consistency towards reality, Use of scientific criteria for the selection of techniques, definitions, methodologies and data sources, Protection of confidentiality. The expected impact is that the new legal arrangements improve the quality, moreover the professional independence of official statistics.

The national legislation and the impact of EU accession requirements have been taken into consideration and evaluated via data points sent to Eurostat and Statistical Management Information System (SMIS+) evaluation. Data Points are statistical indicators and/or quality reports send by NSS to Eurostat. This indicator is calculated by Eurostat and reported to NSS on a yearly basis. SMIS+ is an application developed by Eurostat as important evidence related to the compliance with the EU acquis and CoP. SMIS+ use two sets of criteria to evaluate the compliance levels: data transmission that is calculated by the number of dataset transmitted from Electronic Dataflow Administration and Management Information System (eDAMIS) and completeness of them. The second criteria is for the quality of the official statistical measured by the accuracy, timeliness and punctuality, comparability and coherence.

[3] European Commission (2005)

The internal factors influencing the governance model like human resources management, appropriate budget management, internal governance model and communication, as well as effective and efficient planning processes were evaluated. Human resources management, internal governance model and communication are indicators that were measured using Staff Satisfaction Survey (SSS).

Effective and efficient planning processes have been evaluated by the impact of the new law, related to INSTAT steering body, budget implementation and revenues distribution.

3. Results

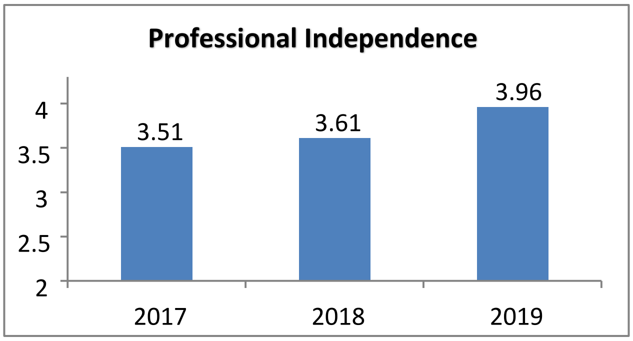
The legal framework is crucial for defining the governance model of the NSS. In Official Statistics new law, mechanisms to collect, process and publish official statistics, are specified. It empowers the statistical agencies in their data collection and production processes. In order to achieve this goal provisions are included to allow the production of official statistics, rights to access confidential data for scientific purposes, right to access and use administrative data, as well as the obligation to the statistical units to provide full and truthful information to INSTAT or other statistical agencies, while respecting statistical confidentiality and privacy.

Under the Law on Official Statistics, INSTAT coordinating activities of NSS include the cooperation with statistical agencies in preparing and implementing the OSP; guidance to the methodologies used for statistical production in compliance with international standards and representation of the NSS in the European Statistical System and international organizations.

Controlling and implementation of statistical quality has been enforced by adding legal provisions regarding the conduct of statistical audits on official statistics producers. It is INSTAT General Director's duty to guide the coordination of NSS and request for statistical audits on statistical production processes within the NSS, when appropriate.

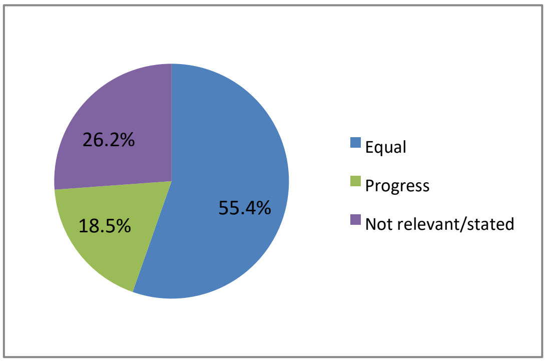
As shown in the Figure 1, the new law and the higher level of hierarchy of the institution has directly the impacted the user perception regarding the professional independence.

Figure 1: Professional Independence, USS 2017-2019



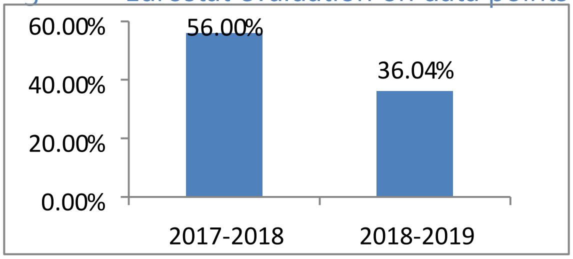
Good governance helps defining priorities in terms of sustainability to determine policies, strategies and action plans necessary to achieve the expected outcomes. Priorities related to official statistics are defined also by the Albania EU accession process. INSTAT, along with statistical agencies, has the duty to produce statistics in a transparent, impartial and timely manner, helping users to judge the performance of the transformation processes towards EU integration (Stabilization Association Agreement (SAA)). As part of the fulfilment of the SAA, and the implementation of the recommendations of the EU for Albania alignment with the acquits under Chapter 18, "Statistics", INSTAT tries to create a culture of statistical development among NSS, monitoring efforts to carry out the obligations of the SAA by the statistical agencies. SMIS+ evaluation over 2017 – 2018, related to the NSS compliance with the EU acquis and CoP, is presented in figure 2.

Figure 2: SMIS+ evaluation 2017-2018



On a four range level of compliance, figure 2 is showing the percentages of improvement made from 2017 to 2018 by NSS.Another indicator presenting the progress done towards reaching the obligation of SAA is the data points transmitted from NSS and published by Eurostat. Figure 3 is presenting the Eurostat evaluation in percentage point.

Figure 3 : Eurostat evaluation on data points



In the work towards reaching INSTAT strategic Objective: "Use of appropriate data sources and instruments to improve the production and quality of indicators and statistical services"[4], over the last years more data has been transmitted to and published by Eurostat. To increase the usage of administrative data for statistical production, INSTAT has been working in the last years to increase the number of memorandums of understanding (MoUs) with administrative data owners as well as research community. The MoUs include not only the cooperation between the parties but also specify the data to be transmitted, frequency and technical provisions. In addition, for the MoUs to be efficient, technical working groups have been established with the main objective to advice and monitor the performance of the activities specified in the agreements. In 2017 – 2019, INSTAT has signed 25 different MoUs, with administrative data owners and universities; 20 of them regarding the administrative data providers and 5 with public and private universities in Albania and abroad. Collaboration with the academic world has also been identified as a priority element, because this interest group helps to improve production standards and use of official statistics in the field of research.

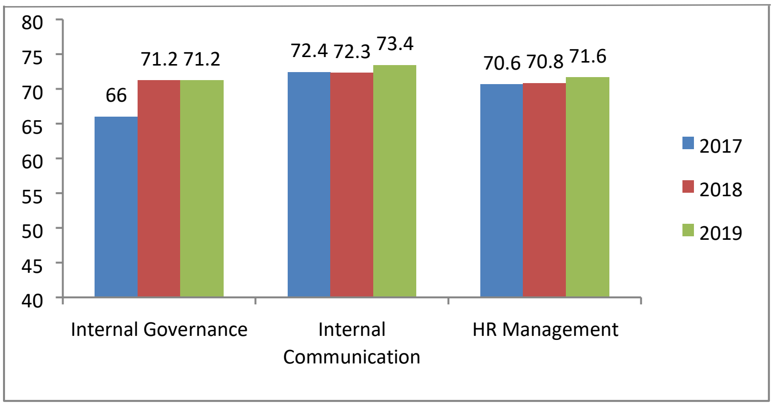
To ensure an appropriate planning, follow up and controlling, clear mechanisms need to be developed supporting the production and dissemination of high quality statistics. These mechanisms need to take into account human resources management, appropriate budget management, internal governance model and communication, as well as effective and efficient planning processes.

Transforming INSTAT in a learning organisation, towards building in house management skills, that enforce teams in reaching goals, is the institution's focus in the last years. Knowledge management is part of INSTAT's training strategy, which is focused in key priorities like quality, administrative data usage for statistical production as well as training on statistical literacy. One of the pillars of this strategy is the organisation of trainings in a unified system, assuring an adequate level of statistical knowledge among official

[4] Development Strategy of INSTAT 2017 – 2030 http://instat.gov.al/media/3986/development-strategyof-instat-2017-2030.pdf

statistics producers. In this regards a Statistical Training Centre is created at INSTAT to increase the statistical capacities of statisticians and statistical literacy among users and different stakeholders, as well as to provide the needed infrastructure for scientific research activities. In this regards, during 2018 INSTAT has organised trainings on the legal and quality management framework for statistical production, as well as basic knowledge in statistical production processes. Over the last three years, progress has been shown in the internal governance factors, based on the SSS findings, figure 4.

Figure 4: SSS 2017-2019



Initiatives to improve internal governance model and communication have been taken from INSTAT. In order to have a more sustainable statistical institution, independent by political interferences and outside factors, the law has clear provisions on the duties of the Director General and Steering Body. The Steering Body is the governing body of INSTAT, composed by the mid and high level management of INSTAT, chaired by the Director General. Its responsibility is to define INSTAT development strategy and propose budget implementation and revenues distribution. The results of SSS are analysed and action plans to improve the level of satisfaction are prepared and monitored. One of the actions taken to improve the satisfaction was the leadership program. This program goal was to improve internal communication and involve staff in the planning and decision-making process.

The new legal framework has increased also the autonomy in managing financial resources. INSTAT can provide statistical services to third parties, performing activities that are outside the scope of OSP. The revenues generated from the provided services can be managed to improve statistical production infrastructure, increase the number of statistics produced based on user's needs, increase staff capacities for NSS and conduct scientific research, studies or analysis. In 2017, one important action, towards improvement of the staff satisfaction, was the increase of the salary level for all INSTAT staff.

4. Discussion and Conclusion

In this paper external and internal factors impacting the governance model have been evaluated towards improving the quality in the NSS. The factors taken into consideration, presents direct impact into the quality of statistical production. The new Law in Official statistics and the new governance of the institution have improved dimensions of quality, such as Professional independence, Coordination and cooperation, Mandate for Data Collection and Access to Data, Adequacy of human and financial Resources.

To achieve the required quality, the NSS should build a model that empowers law, decision making, human resources, communication and the satisfaction of users and staff. Using the model suggested in this paper is concluded that if there is any improvement in one of the above factors, the target quality can be improved.

Qualitative and reliable statistics are very important in the decision-making process at national and local level as well as towards the process of European integration. The enhancement of collaboration and strengthening of NSS will bring direct impact on the production of SDG indicators.

To continue the work of strengthening the NSS, INSTAT should build a developing strategy to support INSTAT and the other statistical agencies in defining strategic objectives for the development of official statistics and synergies among public bodies involved in statistical production or administrating data. In defining this strategy, the results of this paper need to be taken into consideration for development and establishing coordination mechanisms with data users and producers.

References

1. European Commission (2005), “Communication from the Commission to the European Parliament and to the Council on the independence, integrity and accountability of the national and Community statistical authorities, Recommendation on the independence, integrity and accountability of the national and Community statistical authorities”, (COM (2005) 217).
2. Business Process Management 2nd edition by Jeston, John, Nelis, Johan (2008)
3. Law no. 9180, dated 05.02.2004 "official statistics”,
4. Development Strategy of INSTAT 2017 – 2030
5. Law No.17/2018 on Official Statistics

Functional dependence of a National Statistical Office within its respective executive branch. Advantages and disadvantages

Jorge Todesca[1], Hernán Muñoz[2]
National Institute of Statistics and Censuses (INDEC), Argentina

Abstract

The independence of NSOs (National Statistics Offices) can be approached from different angles. There seems to be a broad consensus on statistical methods and processes.

However, a question which is not minor is what kind of articulation between NSOs and their corresponding governments does not lead to isolation and a lack of interest that could eventually affect resources.

In this sense, the main challenge is to establish how a statistical system can strengthen the bonds with the rest of the State in a context of dialogue between peers that will not affect its independence.

In Argentina, INDEC, created by 1968 Law 17622, works as a deconcentrated body within the scope of the Ministry of Treasury, but it was initially within what was then the Planning Secretariat of the Presidency of the Republic.

The diversity of statistical production and the interrelation with different areas of government and sectors of society invite us to reflect on the most adequate modes of articulation for the future development of National Statistical Systems.

1. General considerations

The recent experience of the National Institute of Statistics and Censuses (INDEC) of Argentina has led to deliberation from various perspectives as regards its relationship with the Executive Branch.

Firstly, given the political intervention that took place between 2007 and 2015, it has been a priority in the transformation agenda to establish a legislation that ensures the Institute's complete autonomy and allows updating the available tools for the performance of its functions. Although de Executive Branch has not yet presented a new legislative proposal to Congress, there has been a lot of debate within the Government and there is already a draft law that moves towards autonomy. In terms of governance, it proposes the establishment of an executive management with defined stability and an advisory council composed by government and academic representatives.

[1] Director-General of the National Institute of Statistics and Censuses (INDEC), jtodesca@indec.gob.ar

[2] National Director of Planning, Institutional and International Relations of INDEC, hmunoz@indec.gob.ar

Secondly, even with legal autonomy, the question arises as to the Institute's scope of operation. Considerations on this matter may be generically extended to other national statistical offices (NSO). In our view, there is no doubt that in a presidential system (such is the political organisation of Argentina) NSOs must operate within the scope of the Executive Branch. The creation of a sphere independent from the three branches of Government, as in the case of Mexico, is difficult to implement and is worthy of an analysis that exceeds the scope of these notes.

A direct link with the Executive Branch may give rise to certain degree of tension. The institutional structure of an NSO should therefore include effective coordination powers on this matter.

In other words, having full autonomy without an effective bridge with the Executive Branch entails the risk of isolating the statistical system from government policies. This might eventually lead to a breach of the principle of public policies based on evidence.

To avoid this, there must be a strong institutional will, firmly installed at the political level. Such a remark may be redundant for mature democratic systems, but this is not the case in many countries.

A good coordination between the Executive Branch and the NSO also leads to better chances of meeting the interests of users. In a way, there can always be transactional situations that result in disadvantages for an NSO (and its users) when budget allocations are decided.

The third unavoidable matter is defining to which area of the Executive Branch the NSO should belong. A brief review of the Argentine experience offers certain considerations that may be useful. The law by which INDEC was created was issued in 1968, when the country was under a military administration. The law was therefore not approved by Congress, which was dissolved at the time, but by the President of the Republic, invested with legislative powers by the founding acts of the military after the 1966 coup. It is a rather strange case. The law abolished all previous fragmented legislation and created an institute under rather modern canons, delegating regulatory powers over official national statistics and creating a National Statistical System. Its creation was led by a prestigious statistician at the time, but it was not accompanied by the public debates that these foundational actions usually require. Neither did the press register the sanction of the law or the complementary regulations that followed. Although the law did not give autonomy to the Institute, INDEC functioned with few disturbances until the political intervention that took place during the 2007-2015 period. Until then, the Institute had been professionally respected and had become one of the most prestigious institutions in the country. It also participated actively in the international statistical community.

At the time of its creation, and for several years, the Institute worked within the scope of the Presidency, under the National Development Council (CONADE, for its Spanish acronym), an institution created in accordance with the developmentalist ideas of the time. It was the seedbed for the first INDEC teams, even though the political changes did not encourage a connection between official statistics and government policies.

In the 1990s, after several political changes in the Argentine National Administration, INDEC began to operate under the Ministry of Economy or Ministry of Treasury, depending on the subsequent names it was given.

This dependence was accompanied by the reinforcement of economic statistics. Both national accounts and balance of payments, which until then were the responsibility of the Central Bank, were placed under INDEC's responsibility.

Seen under a current perspective, although social statistics have had a strong presence and continuity, there have not been strengthening actions in this area on a systemic level. The Institute's pillar has always been the continuous household survey, to which several other periodic sectoral surveys have been added, in cooperation with other areas of government in fields such as health risks, child labour and victimisation.

In a way, experience shows that the traditional dependence of the NSO to economic areas of government may lead to an asymmetric view as regards the current field range of official statistics.

We may take as an example the framework of well-being indicators adopted by the Organisation for Economic Cooperation and Development (OECD) and fostered beyond the Organisation. Accordingly, several studies have revealed the need to search for economic performance and social progress indicators beyond the classic measurement of Gross Domestic Product (GDP). This insight is reinforced by the fundamental principles of the Stiglitz-Sen-Fitoussi Commission[3] work. There are several other studies that look into the possibilities opened by measurements that go beyond the GDP.

Based on INDEC's past and present administrative status, it can be said that an NSO should operate under a body that encourages a broad view of public policies, such as a Ministry of Planning, the Prime Minister's office or, as would be the case in Argentina, the Chief of Cabinet of Ministers' office.

It is not only a matter of working under the scope of the Executive, but also under a body where the NSO can establish a better connection with all Government policymakers and actively participate, to channel the users' interest when designing short or medium-term working programmes.

A brief account of the INDEC's institutional history follows.

[3] See: https://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report.

INDEC's institutional history

The National Institute of Statistics and Censuses (INDEC) of Argentina was created in 1968 by Law 17622. The legal framework was later expanded by executive orders 3110/1970 and 1831/1993 and INDEC Provision 176/1999. In its capacity as governing body of the National Statistical System (NSS) of Argentina, INDEC regulates on national statistical matters and manages all official statistics carried out in the country. The Institute is a public deconcentrated body, of a technical nature, within the scope of the Argentine Ministry of Treasury, and its Director-General is at the level of a Secretary of State. Due to Argentina's presidential system of government, the Ministry of Treasury is directly under the Nation’s President, as the Constitution establishes that all executive power falls on the President.

Regarding the deconcentrated status, this is a mechanism in which ministries delegate powers to other entities within their scope. It must be noted that this status is different from the “decentralised” status. Although the State delegates powers to a deconcentrated body, the latter does not have its own legal personality, nor can it undertake obligations. Decentralised entities do have these characteristics.

The current legal framework establishes, among other matters, the following responsibilities for INDEC:

- to implement the statistical policy of the Argentine State;
- to structure and rule the NSS;
- to design statistical methodologies;
- to organise and manage statistical infrastructure related operations; and
- to produce basic statistical indicators and social, economic and demographic information.

The NSS is legally centralised and functionally decentralised. This means that INDEC has regulatory authority over the central statistical services of the National Government and the peripheral services, such as the Provincial Statistical Offices.

Until early 2007, the technical independence of the Institute had been respected by all National Public Administrations in matters such as the definition of statistical methodologies, statistical production processes and the publication of technical reports. For most of its more than 50 years of existence, INDEC was considered to be a highly professionalised entity with great technical capacity and a sound credibility of its statistical production. The Institute also had an active international involvement in various specific forums and exchange programs. These variables gave the Institute a de facto protection of its professional independence.

The political intervention that took place between 2007 and 2015, widely reported both nationally and internationally (TODESCA and MUÑOZ, 2017), resulted in an institutional deterioration process which included the replacement of several directors and the arbitrary displacement of many specialised statistical agents, who either resigned, were dismissed, or, in many cases, continued their work in a hostile environment if they opposed the regime.

The discretionary changes in statistical methodology and procedures and, especially, the spurious statistical indicators published resulted in a wide-spread mistrust of the Institute's production. An evident scepticism grew within the Institute's users, who began to use alternative indicators to approach the social and economic reality. Additionally, INDEC lost its statistical leadership within the Argentine State and its status in the NSS at all levels.

In December 2015, at the beginning of the country's current Public Administration, the scenario at INDEC was of great distress due mainly to three aspects: low credibility, institutional status and lack of leadership over the NSS. In order to reverse this situation, INDEC management team focused on the following political and administrative tools:

- Decree 55/2016. Declaration of emergency in the National Statistical System.
- Request to adhere to the Recommendation of the OECD Council on Good Statistical Practice.
- Management team renewal.
- Granting a higher hierarchical status to the Director-General of the Institute (from Undersecretary of State to Secretary of State).

A higher hierarchical status enabled, on one hand, direct access to the highest National public administration levels. On the other hand, it gave the Institute the capacity to issue certain administrative rules and documents without approval of a superior authority. Thus, as a Secretary of State, INDEC's Director-General can issue resolutions without an intermediate entity[4]. Before the hierarchical change, the Institute's Director-General could only directly issue provisions[5]. Both resolutions and provisions are ranked below executive orders (official orders that are the result of a decision of the National Executive Power and can only be signed by the President).

As mentioned in the brief historical account, after its creation in 1968 under the scope of the Secretariat for the CONADE, the Institute was transferred to the National Planning Secretariat, and, later, to the Ministry of

[4] Administrative decisions issued by ministers, ministers' or presidents' secretaries, or other officials authorised to do so and which, depending on the specific topic, could be effective both within and outside the respective organisation.

[5] Decisions from lower-level administrative authorities (such as an undersecretary, the head of a decentralised organisation or a director-general) on issues related to their respective organisational sphere.

Economy (in its different denominations)[6]. Within this Ministry, it became an Undersecretariat under the Economic Policy Secretariat until 2009. Later, INDEC was subordinated directly to the Minister of Economy. Although the Institute was under the Ministry of Economy until 2007, there is no evidence of important interference or relevant pressure from the superior entity to the Institute.

In its beginnings, and during most of the 1970s, INDEC was mainly focused on socio-demographic statistics, such as population, employment, unemployment and poverty. Since the 1980s, the Institute progressively increased its economic statistics production, a strategic guideline which was reinforced in the early 1990s when INDEC began to produce both the country's national and international accounts. Setting aside professional independence, the fact that INDEC performs its activities under the Ministry of Treasury has its advantages and disadvantages in connection with its administrative and institutional dynamics. Among the advantages, the following may be mentioned:

1. It enables access to budget resources, in terms of quantity, timeliness and specific characteristics of requirements, given that the Ministry of Treasury drafts the National Budget.
2. It allows close interaction between Ministry and INDEC officials, improving the definition of priorities, the characteristics of statistical operations and the potential added value for the definition, monitoring and assessment of economic policy. In this sense, the Ministry of Treasury is the main interested party in INDEC's statistical production.
3. It helps establish communication channels to better explain statistical matters to policymakers within the Ministry.

Thus, the vertical integration between INDEC's statistical production and the Ministry of Treasury substantially reduces asymmetries in information between the entity that produces statistical indicators and its main user within the National Public Administration. Moreover, being able to access, with the Ministry's endorsement, the available financial resources for statistical projects and operations could increase the Institute's budgetary efficiency and improve its resource management.

Therefore, the advantages could be summarised as follows:

- better access to the entity that drafts the budget;
- articulating demands with the main client;

[6] This Ministry was named, successively: "Ministry of Economy and Labour" (1966-1971), "Ministry of Treasury and Finance" (1971-1973), "Ministry of Economy" (1973-1991), "Ministry of Economy, Public Works and Services" (1991-1999), "Ministry of Economy" (1999-2002), "Ministry of Economy and Production" (2002-2008), "Ministry of Economy and Public Finance" (2008-15), "Ministry of Treasury and Public Finance" (2015-16) and since then "Ministry of Treasury".

- better use of public statistics.

Nevertheless, under certain circumstances, these three advantages could become a disadvantage to a National Statistical Office, specifically, as described by Seltzer (1994)[7], when the political power intends to institutionally undermine an NSO. Even more so if the superior entity is the most affected by the uncomfortable truths told by the NSO in terms of social or economic performance.

According to the current governance structure, INDEC acts under the Ministry of Treasury, which in turn operates under the authority of the Nation's President. This means that INDEC tells "uncomfortable truths" that reflect the performance of many of its direct superior's policies. Thus, the Institute assumes an internal and external evaluating role, since, while the President of the Republic is accountable to society, the Minister of Treasury is accountable to the President. The structure described above could generate several conflicts among the parties involved, according to the so called "agency theory". It must be noted that this is not currently the case.

In this sense, functional dependence from ministries in charge of economic policies implies several eventual drawbacks for an NSO, such as:

1. Intrusion of political officials who may want to implement methodological changes or ask for specific revisions of the NSO's statistical data.
2. Pressure through budgetary subordination.
3. Biases both in the definition of priorities and of the main features of certain statistical operations, with the risk of having the ministry's interests privileged over statistical objectives.
4. Damage to user trust in the NSO's statistical production due to the direct dependence from the ministry.

Beside the above-mentioned disadvantages, there are others that also impact negatively on INDEC's degrees of freedom as regards its operational and administrative management. A deconcentrated entity has administrative disadvantages that hinder its daily activities due to bureaucratic and administrative procedures that require authorisation by the superior entity, such as:

---

[7] According to Seltzer (1994), there are several ways to undermine the professional integrity and credibility of statistical agencies. Generally, the specific threats to statistics affect a wide variety of dimensions, such as: 1) the mission of the statistical service, 2) the financial resources and controls, 3) staff, 4) suppression or modification of statistical fields or series, 5) definitions, concepts and methodology, 6) terms and nomenclature, 7) altering specific numbers, 8) the extent and timing of the release of data, 9) threats to data confidentiality, 10) use of agency for political analysis or other political uses, and 11) active campaign to discredit statistical outputs, methods or staff.

- procedures related to missions abroad carried out by INDEC officials (for specific institutional or technical reasons);
- human resources management, especially in connection with contracting (which, in turn, negatively affects planning, development and performance of statistical operations, usual tasks and projects); and
- appointment and status of directors and coordinators of second hierarchical level.

In conclusion, although INDEC currently still suffers from important normative deficits in de jure protection as regards many of the disadvantages here presented, there is a de facto protection supported by the political will of the current National Public Administration, which guarantees the efficient provision of the country's official statistical service. However, there is still no legal instrument to ensure that the herein mentioned disadvantages will not arise during a future Administration. It is therefore understood that persevering in the promotion of de jure and de facto actions to strengthen, guarantee and enhance the Institute's technical, operational and administrative independence is a virtuous objective for INDEC to carry out its activities in a more efficient and effective way and in line with the current international statistical good practice principles and recommendations.

References

1. DARGENT, E.; LOTTA, G.; MEJÍA, J. A.; MONCADA, G. (2018). "Who wants to know?: The Political Economy of Statistical Capacity in Latin America". Inter-American Development Bank (IDB).
2. SELTZER, W. (1994). "Politics and Statistics: Independence, Dependence or Interaction?" . Working Papers series N° 6. United Nations, New York.
3. STIGLITZ, J.; SEN, A.; FITOUSSI, J.P. (2009). "Report by the Commission on the Measurement of Economic Performance and Social Progress". Commission on the Measurement of Economic Performance and Social Progress. Paris, France.
4. TODESCA, J. (2018). "The Challenges to the Independence of Official Statistics: The case of Argentina". IAOS-OECD Special Meeting. Paris, France. Recovered: https://www.iaosisi.org/images/IAOS201719/2018Conference/SpecialMeeting/Session1_2_Jorg e_Todesca18Sept.pdf.
5. TODESCA, J.; MUÑOZ, H. (2017). "The main challenges related to the rebuilding of INDEC in the Republic of Argentina", ISI2017. Marrakech.
6. UN General Assembly, Res. 68/261. "Fundamental Principles of Official Statistics". A/RES/68/261 (New York, January 29, 2014). Recovered: https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf.

Influence of the digital era and globalization on the functioning of national statistical societies – Hungarian experiences

Eva Laczka, PhD
Hungarian Statistical Association, president

Abstract

The aim, mission of national statistical societies is to help the theoretical, practical development of statistics, create a forum to discuss theoretical and practical issues of statistics, take a stand on statistical methodological questions, inform about current statistical issues of public interest.

The goals of the national statistical societies that have emerged since the first half of the nineteenth century were quite similar, they are in many respect perennial goals. Economic, social processes, the development of sciences and technology have continuously changed over the decades, and have influenced the functioning of statistics and national statistical societies. The decade, the years after the turn of the millennium have brought changes more significant than ever. Information society, the new communication channels, the modern IT and telecommunication systems have created situations, new phenomena not known earlier in the contacts between people. Nowadays one of the greatest challenges for statisticians is the observation of the new phenomena, processes of our globalising world, the correct interpretation of integrated processes. Besides observing the new phenomena, statistics has the possibility to apply new methodologies, technology, technical tools, nowadays the use of new methods and technologies has become a key necessity. The new phenomena, the new requirements have influence on the scope of the statistical fields to be observed, the "new definition" of statistics, statistical methodology, the ways of communicating data strongly influence the issues of statistical culture as well.

The lack of trust in statistics and statistical data gave birth to professional ethics in statistics, the need to discuss ethical issues too. The above mentioned factors affect official statistics, researchers and all the stakeholders engaged in statistics, and they affect – particularly – national statistical societies. The purpose of the paper is to evaluate to what extent the situation, the role of national statistical societies differs from the situation and role of other actors of statistics. What kind of tasks, "stronger prerogatives "could be attributed to national and international statistical societies, international organisations compared to official statistics and other stakeholders. What kind of cooperation should be established between stakeholders in statistics nowadays?

Keywords
Globalization; IT development; statistical culture; ethics; cooperation

1. Introduction

The roots of statistics date back centuries, millenniums, its emergence can be linked to the formation of class societies, states. Statistics was used first in taxation and strategy of states. Population censuses played an outstanding role among the sources, data were collected on the population of the states and the economic, social processes that were considered important at that time.

Due to the development of sciences and technology, official statistics was created in the 18th century (as a component of modern national state), followed by the successive creation of national statistical societies. European statistical offices and national statistical societies were established one after the other in the second half of the 19th century. The Hungarian Royal Statistical Office was also established in that period (1871), the Hungarian Statistical Association was founded only later, in 1922.

In the seventies in the 20th century IT development had a strong influence on statistics, register-based statistics appeared including the system of data protection too. The European Statistical System was also established in that period as well.

The 2010s have brought unprecedented changes in statistics; globalisation and information explosion have created a fundamentally new situation for statisticians and users of statistics.

2. Methodology

2.1. Current challenges

The goals of statistics that have emerged since the first half of the nineteenth century were quite similar, they are in many respect "evergreen" goals. Economic, social processes, the development of sciences and technology have continuously changed during the decades,and have influenced the functioning of statistics and national statistical societies.

The decade, the years after the turn of the millennium have brought changes more significant than ever. Information society, the new communication channels, the modern IT and telecommunication systems have created **new situations, phenomena not known earlier** in the contacts between people. Nowadays one of the greatest challenges for statisticians is the observation of the new phenomena, processes of our globalising world, the correct interpretation of integrated processes.

Besides observing the new phenomena, statistics has the possibility to apply new methodologies, technology, technical tools, nowadays the use of new methods and technologies has become a key necessity in statistics. The new phenomena, the new requirements have influence on the scope of the statistics to be observed. The **new "definition" of statistics, statistical methodology to be applied, the ways of communicating data strongly influence the issues of statistical culture as well.**

The **decline of trust in statistics** and statistical data gave birth to **professional ethics in statistics**, the need to discuss ethical issues, issues statements has increased. The above mentioned factors affect official statistics, researchers, and all the stakeholders engaged in statistics but they affect – particularly – the operation of national statistical societies.

According to the **new vision of statistics** in the future statistics could be produced more outside the official statistical service as well (such as researchers, universities, academic sphere, but by NGOs too. This requires nevertheless more than traditional cooperation; the methodology of the statistics produced should be harmonized, integrated, the standards of statistical methodology and ethics should be respected and enforced. It is obvious that the „validation" of statistics produced in the different spheres represents an immense task both for official statistics and national statistical societies too.

2.2. The actors of statistics

Nowadays the most important actors in statistics are the national statistical offices and the national statistical societies. Despite similarities, the tasks, challenges, operation of the most important statistical actors differ.

2.2.1. The national statistical offices and their tasks

The number one „producers" of statistics are the national statistical offices, the institutions of national official statistics. As it is the case for many other institutions, the task of the Hungarian Central Statistical Office (KSH) is to elaborate statistical concepts, methods, conduct the data collections stipulated in the national statistical system (with particular regard to population censuses), and process the data. It considers as its task to prepare analyses and provide statistical information for public bodies, social institutions, parties, local governments, scientific research and public opinion.

2.2.2. The national statistical societies and their tasks

The tasks of national statistical societies are more diversified. The aim, mission of the Hungarian Statistical Association is to help the theoretical, practical development of statistics, create a forum to discuss theoretical and practical issues of statistics, take a stand on statistical methodological issues, inform about current statistical issues of public interest.

The fact that the tasks of national statistical societies differ from those of the official statistical services arises from the more diverse composition of national statistical societies. Beyond the institutions of the official statistical service, the actors of the academic sphere, the university campuses dealing with statistics, enterprises using statistics for their operation participate also in the work of national statistical societies. What is important is that in the national statistical societies we also find users in the broad sense of the term, citizens, public opinion interested in statistics. Due to the composition of national statistical societies, the international relations of the national statistical societies are also more varied, which could be the source of many advantages. What is also important: they have greater „freedom" in freedom of expression.

2.2.3. International relations in statistics

The birth of statistics was followed by the emergence of international statistical organisations. We have to mention among the firsts the **International Statistical Institute** (ISI), which was founded in 1885 after several years of preparation.

One of the most important fora of official statistics is the **UN Statistical Commission** established in 1947, which is basically a body of official statistics, but has begun in the past years to deal more and more frequently with the issues of the cooperation between official statistics and national statistical societies as well.

As European citizen, I would mention the federation created by the European national statistical societies, the **Federation of European National Statistical Societies (FENStatS)**. FENStatS was established in 2011, currently having 24 members (the Hungarian Statistical Association joined the Federation in 2013). The aim of the FENStatS is to promote mutual communication, co-operation and statistical research in Europe and to develop, relations of statistics to society, relations to European institutes, and particularly has the scope to support the diffusion of statistical education in Europe.

In Europe we can find significant number of cooperation activities established on the basis of other motivations. Such a cooperation is – for example – the joint action of the **German-speaking European national statistical societies,** or the cooperation of the national statistical societies in the Central European region (**Visegrád** national statistical societies - V7). The V7 was established in 2005 in Budapest by the Hungarian Statistical Association. The members of the **V7** cooperation are the national statistical societies of Austria, the Czech Republic, Hungary, Romania, Slovakia, Slovenia and Poland (it should not be mixed up with the V4 political formation established in the 2010s by Hungary, Slovakia, the Czech Republic and Poland).

The V7 statistical cooperation established in 2005 is a professional forum. The aim of the V7 is to facilitate scientific progress and theoretical, practical development of statistics in the region. Provide a regional forum for the discussion of statistical issues and take up a position in professional matters. Maintain contact with the national statistical societies in the region as well as offer support to nascent statistical societies in the region and contribute to the elaboration of the Codes of professional statistical ethics in the region.

The statistical associations mentioned above represent different „geographical levels" of cooperation, they constitute in a certain sense regional cooperation. I am convinced that each of the cooperation has its mission and tasks, in the majority of the cases they are facing quite similar specificities, same problems and motivations.

2.3. The feature of changes

The significant changes that have taken place in the past years concern the whole statistical process, they influence the aims, tasks, methodology of statistics, but they change the functioning of the statistical actors as well.

Let us think about the need of **standardising, harmonising statistical concepts, definitions,** and the global treatment of statistical systems.

The **„wealth of information, data"** caused by information explosion urges statisticians to replace "traditional data collections" by new data sources (Big Data, AI, remote sensing data, etc.). While new data sources allow to produce statistical data, information more quickly and in a cheaper way, the new procedures decrease respondent burden as well. Despite their advantages, methodological challenges are much greater than previously. The focus on the methodology of data collections is shifted towards the use of mathematical statistical methods, modelling. Nowadays statistical work requires a greater integration of mathematics and other sciences.

Similar considerations apply to the **processing of statistical data** as well. Due to the explosive development of IT, innumerable IT tools are at the disposal of statisticians. Nowadays IT developments allow „tailor-made" solutions for statistics, statistical data processing can be governed by statisticians. In Hungary a good example is EAR (Unified Data Processing System) developed by the Hungarian Statistical Office. EAR speeds up statistical data processing, improves data quality, it secures a better documented validation of data processing, and allows even automatic reprocessing of the statistical data. What is important is that statistical data processing belongs to the competence of statisticians, it integrates statistical and IT work for statisticians.

The elaboration, introduction of new methodologies, the use of new IT tools require the reconsideration of **statistical knowledge** too. Not to speak about the fact that **statistical thinking** is nowadays needed in all sciences, it is an indispensable element of the operating culture.

It is of course obvious that statistical knowledge, culture have different challenges for the „producers" of statistics and for users of statistics in broader terms too.

2.4. Reputation of statistics, statistical culture

The credibility of statistical data has been questioned nearly since the birth of statistics. There is no other science, profession surrounded by so many jokes, anecdotes, and frequently even serious criticism. What is the reason of the constantly present lack of trust?

I am convinced that among the reasons we should basically highlight two considerations; on the one hand the presentation of **data are often linked to interests,** but it is also frequent that users **do not possess the necessary statistical knowledge.** Users of statistical data – often – use statistical data, information according to their interests and they manipulate the data.

There are numerous examples which show that **users of statistical data are not aware of** (or do not take into consideration) the exact **meaning of the data, definitions,** the relations between the pieces of information (**they do not know the metadata**). It is frequent that users are not aware of the data sources and the way data are produced, they do not know for what purpose and how data can be used.

The question arises how nowadays practising statisticians and statistics teachers are able to cope with the stereotypes, how they can convince users, public opinion that statistics can be credible as well. This **requires** of course **a thorough statistical knowledge and enhance of statistical culture.**

2.5. The grouping of users and their statistical culture

There are of course many examples for the grouping of users. The most frequent solution is to classify users into groups according to the institution they belong to. I guess, when we are dealing with the statistical culture (statistical knowledge) of users better results can be reached if we classify users according to the way they use data.

I personally prefer the grouping I heard years ago from Danish statisticians. The colleagues from Denmark had defined three user groups: *„tourists", „farmers"* and *„miners".*

„Tourists" are the users who do not use statistical data on a regular basis, in certain cases they may be interested in one data, indicator, and they try to find it via internet. „Farmers" are the users who use data for their operation on a regular basis. Finally, „miners" are the users who analyse deeply a given subject, or some specific field of statistics regularly.

A certain level of statistical knowledge (statistical culture) is needed in the case of all three groups of users. It is well-known that – in the majority of the cases – „tourists" have the poorest statistical knowledge, culture while the knowledge of the „miners" is – in most of cases – adequate. The improvement of statistical knowledge, culture is an important issue for all three groups, but requires very different tools. While in the case of „miners" and „farmers" tailor-made courses, trainings can lead to good results, other tools (actions) should be designed, used for „tourists" and for public opinion.

3. Results

Some examples

For a few years ago, the Hungarian Statistical Association has been sending quarterly electronic **Newsletter** to all the members of the Statistical Association – while the news are available on its homepage, some of the presentations are available on You Tube too.

The Hungarian Statistical Association has experienced – in the past 5 years – that one of the most efficient tools to shape public opinion is to organise **statistical days in an open-air museum** (Szentendre). During the two days programmes about 2500 visitors of the open-air museum attended the statistical programmes (filled quizzes, participated in different games, learned about statistics, data, about use of data a lot).

The Board of Ethics of the Hungarian Statistical Association has also established a new tool to enhance statistical culture; namely, **they analyse the statistical language published in the written and electronic media**. The purpose of the analysis is to evaluate the anomalies in the use of statistical/economic specialized terminology, including the bad usage of words due to lack of knowledge, but also the intentional „misuse" (manipulation of data).

The primary task of the Board of Ethics of the Hungarian Statistical Association is of course to **elaborate the code of ethics** for statistics, and to monitor ethical conduct and practice. The compilation of the Code of Ethics of the Hungarian Statistical Association was initiated by the Association on the basis of the request of the Statistical Committee of the Hungarian Academy of Sciences. The Code was adopted in 2006 by the General Assembly of the Association. Although the elaboration of the Code of ethics itself was already a huge task and challenge, the real difficulty for the Board arose during the application of the Code of ethics. The assessment of the issues, complaints referred to the Board was not followed by an objective, transparent discussion on the given case, the criticism had serious negative consequences that could not yet be overcome. This shows that there are still measures to be taken in the field of statistical culture.

4. Discussion and Conclusion

The tasks ahead of national statistical societies, international organisations are the followings.

At national level I consider that the national statistical societies should play a key role

- in **developing the vision for national statistics** renewed under the influence of globalisation and the digital era, and in the application of it, they should ensure forum for the discussions,
- play a key role **in integration, harmonisation of up-to-date statistics** emerging in different fields, in strengthening of the communication between actors,
- national statistical societies must play a key role **in getting acquainted with and disseminating good practices** used in statistics,
- should play an outstanding role **in the development of statistical culture** and operating culture, in the development of new tools, the adaptation of good practices,
- They must exploiting the advantages of the digital world, **enhance the communication** between statisticians and users making the operation even more effective.
- The Hungarian Statistical Association broadly agrees with the **concept, vision, paradigm shift**, as formulated at the NTTS 2019 conference held in Brussels in March 2019.

At international level I guess the international statistical institutes, federations should be more open than before, the **information flow should become quicker and more effective**. The international organisations, federations – as the highest level of hierarchy – should give stronger support to the national societies, especially in the case of sensitive, difficult issues, statements.

References

1. Magyar Statisztikai Társaság (Hungarian Statistical Association). (2016). *Fényes Elektől az Európai Statisztikai Rendszerekig* (From Elek Fényes to the European Statistical Systems). Budapest: Xerox Kiadó.
2. Laczka, É. (2015). Hazai statisztikánk a nemzetközi térben (National statistics in the international space). *Statisztikai Szemle (Statistical Review)*, 90(11-12), 1052-1058.
3. Eurostat, European Statistics Code of Practice (2018), Luxembourg: Publication Office of the European Union
4. www.isi-web.org

Individuals using the internet for ordering goods or services related to selected digital and economic development level indicators: Cluster analysis of European countries

Ksenija Dumičić[1], Zagorka Lozanov-Crvenković[2], Emilija Nikolić Đorić [3], Blagica Novkovska[4]

[1] Department of Statistics, University of Zagreb Faculty of Economics and Business, Croatia & Croatian Statistical Association, Zagreb, Croatia

[2] Department of Mathematics and Informatics, Faculty of Sciences, Novi Sad University & Statistical Society of Vojvodina, Novi Sad, Serbia

[3] Department of Agricultural Economics and Rural Sociology, Faculty of Agriculture, Novi Sad University, Novi Sad, Serbia

[4] University of Tourism and Management, Faculty of Economics, Skopje, R. of North Macedonia

Abstract

With a purpose to evaluate the possible role of statistical societies in digitalisation and globalisation era, selected "digital skills – related" and "globalisation" variables are put into the focus of the desk research, hoping that conclusions might be stimulating for statistical societies' enhanced support to "digital skills – related" statistical education, knowledge, literacy and skills. Regarding this, possibilities of national statistical societies' and Federation's of National Statistical Societies (FENStatS, 2019) a more intensive standardized communication is suggested. The goal of the desk research was to explain one digital skills related variable, with three others. In this research, data for Percentage of individuals aged 16 to 74, who use the internet for ordering goods or services ($Y_{2017IntOrderGoods}$), related to three development indicators, as independent variables, for 31 European countries for 2017, has been explored. Among several indicators influencing the main variable under study positively, Percentage of individuals aged 16-74, who have basic or above basic overall digital skills ($X3_{\_DigitalSkill}$), showed to have the strongest positive correlation with it. Descriptive summary statistics, linear trend modelling, correlation, regression and hierarchical cluster analysis were used. Analysis included 27 countries of the EU-28, plus four EU candidates, Montenegro, FYR of Macedonia, Serbia and Turkey. The estimated linear trend for the EU-28 countries shown that Percentage of individuals aged 16 to 74, who use the internet for ordering goods or services increased with the slope of 2.73 percentage points yearly. After discovering quite strong positive correlations between the dependent and each of independent variables, a two multiple linear regression models were fully developed, both including $X3_{\_DigitalSkill}$, as the independent variable. Both models show the average increase in the regression value to be statistically highly significant for each of the regressors considered. After hierarchical clustering, the Scandinavian

countries with Germany and United Kingdom, all clustered within the separate cluster, as the absolute European leaders regarding the Digital Society indicator and Percentage of persons who order goods and services online, while the South-East European (SEE) countries, being similar, were clustered, as those at the end of the queue. Here recognized differently developed clusters of countries, with different percentages of individuals for $Y_{IntOrderGoods}$, and $X3_{\_DigitalSkill}$, variables taken as examples for digitalisation and globalisation, have also different communication habits of their NSAs. Activities of the NSAs should be studied with the goal to standardize their communication, to be better fitted into the global digitalised World, by communicating over the common platform, fulfilling their activities' goals.

Keywords
Globalisation; Digital Skills; Hierarchical clustering; Statistical societies; FENStatS

1. Introduction

Purpose of this study is to recognize statistical societies' role and acting opportunities in the era of digitalisation and globalisation, based on digital related variables research findings. Since the overall present digitalization initiates changes in lives of people in all the societies and creates a growth of economies in countries all over the World, it is considered as the natural component of our reality. In this research, selected indicators, which influence the main variable under study, Percentage of individuals who use the internet for ordering goods or services ($Y_{2017IntOrderGoods}$), which increased for EU-28 from 30% in 2007 to 60% in 2018, were analysed. Among several considered, one Digital Society related indicator, called Percentage of individuals who have basic or above basic overall digital skills ($X3_{\_DigitalSkill}$), showed to be positively correlated, with the strongest positive influence on the main variable under study. Three indicators, with fully available data, are included for correlation and regression analysis: GDP per capita in PPS, Index, EU-28=100; Percentage of households who have internet access at home, for the population aged 16 to 74; and Percentage of individuals aged 16-74, who have basic or above basic overall digital skills. Since the descriptive exploration found that the Luxembourg's Gross Domestic Product (GDP) per capita in PPS, $X1_{\_GDPpcPPS}$, has been an extremely high outlier, the further analyses with 31 countries for 2017 were performed: the 27 countries of EU-28, Montenegro, FYR of Macedonia, Serbia and Turkey. The authors studied mutual interaction of the variables with correlation, regression and cluster analysis in 2017. Dumičić, Žmuk & Novkovska (2017) performed regression analysis of e-commerce, focusing the selected EU candidates and the EU countries. Further, Dumičić, Žmuk & Mihajlović (2017) executed profile analysis of clustered European countries constructed using a number of indicators impacting the e-commerce for individuals. Dumičić, Skoko Bonić & Žmuk (2018) conducted statistical analysis of nine development indicators' impacting the Internet purchases by individuals, defined as a share of individuals who made a purchase by using the Internet in the last 12 months in the total population of a country for 2013. Further, variety of development indicators influences special forms of internet purchases were analysed, too. So, online banking analysis, given in Dumičić, Čeh Časni & Palić (2015), includes variety of multivariate analysis, while development indicators impacts on online booking for travel and accommodation is presented in Žmuk, Dumičić & Mihajlović (2014), and Dumičić, Žmuk & Čeh Časni (2015), showing separate clusters of European countries. Žmuk & Mihajlović (2018), paid attention to position of the Western Balkan countries within Europe regarding online booking influenced by economic and digital development level indicators. Nagy (2016) explored e-commerce on Hungarian market, and Nagy (2017) compared Hungarian and Ukrainian digital economy and society. A number of visualised digital economy impacts of various aspects of customer behaviour is given in Consumer Barometer, as given at CB (2017).

In this paper, a trend analysis for $Y_{2017IntOrderGoods}$, over the period 2007 to 2018, follows. The added indicators have been as follows: $X1_{\_GDPpcPPS}$, Level of internet access for households, given as Percentage of households who have internet access at home, for the population aged 16 to 74, and ($X3_{\_DigitalSkill}$). The research hypotheses state, firstly, that a positive correlation exists between $Y_{2017IntOrderGoods}$ and each of three included explanatory indicators, so that clear clusters of countries could be created; and there are increased motivation and opportunities for statistical societies to act in digitalised and globalised World by improved communication when promoting statistical and digital related improvements in education, knowledge, literacy and skills. Performed statistical data description, performed correlation, regression and cluster analysis, helped testing these hypothesis.

2. Methodology

After investigation on the recent official data availability, Eurostat (2019), the following indicators have been analysed:

- $Y_{2017IntOrderGoods}$ - Percentage of individuals aged 16-74, who use the internet for ordering goods or services, being Digital Society related indicator and the main variable of the interest in this study;
- $X1_{\_GDPpcPPS}$ - GDP per capita in PPS, Index, EU-28 = 100;
- $X2_{\_AccHome}$ - Percentage of households who have internet access at home, for the population aged 16 to 74, also called Level of internet access for households; and

- $X3_{\_DigitalSkill}$, Percentage of individuals aged 16-74, who have basic or above basic overall digital skills.

Firstly, an Ordinary Least Squares (OLS) liner trend model for the variable $Y_{2017IntOrderGoods}$ for the EU28 area over the period 2007 to 2018, and short term forecasts for 2019 and 2020 were estimated.

Further, all four considered variables are described using Descriptive Statistics for 31 countries, as shown in Table 1. Since, in exploratory data analysis the Luxembourg's data for $X1_{\_GDPpcPPS}$ appeared to be a serious outlier, with z= 3.84, it must be removed from further analysis. The EU-28 countries, without Luxembourg, with Montenegro, FYR of Macedonia, Serbia and Turkey, in 2017 proceeded for the further analysis.

Multiple linear regression (MLR) analysis, using the software *Megastat* and *EViews,* was used, Gujaraty & Porter (2010). The population model is given in (1):

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{j,i} + e_{i,} \qquad i = 1,2, \ldots\ldots, n \tag{1}$$

Regression parameters are estimated using the OLS method. The model with estimated parameters is given in (2):

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_{j,i}\,, \qquad i = 1,2, \ldots\ldots, n \tag{2}$$

And, finally, for $Y_{2017IntOrderGoods}$, and three explanatory variables, $X1_{\_GDPpcPPS}$, $X2_{\_AccHome}$ and $X3_{\_DigitalSkill}$, a hierarchical cluster analysis was performed.

3. Results

OLS liner trend model for the variable $Y_{2017IntOrderGoods}$ for the EU-28 area over the period 2007 to 2018, and short term forecasts for 2019 and 2020 in Figure 1 are illustrated.

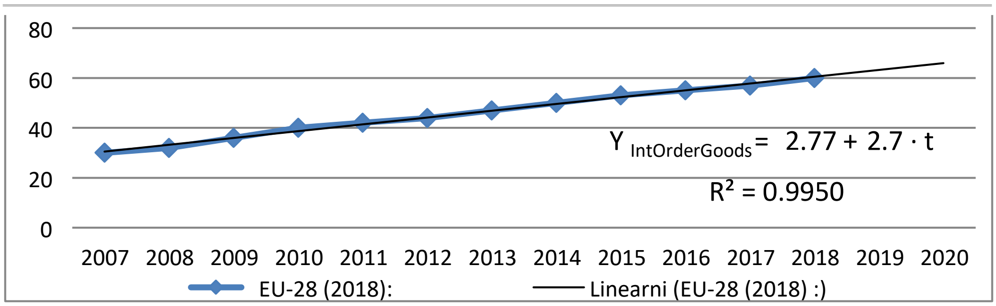


Figure 1. $Y_{IntOrderGoods}$, *EU-28, period 2007-2018 and forecasts for 2019 and 2020*
Source: Eurostat, Authors' creation.

Table 1. *Descriptive statistics, for n=31 selected European countries in 2017*

	Y2017IntOrderGoods (percentage)	X1_GDPpcPPS (Index, EU-28 = 100)	X2_AccHome (percentages)	X3_DigitalSkill (percentages)
Count	31	31	31	31
Mean	47.48	87.10	82.61	53.61
Standard Error	3.78	5.90	1.50	2.46
Standard Deviation	21.05	32.84	8.38	13.70
Coeff. of variation=	44.33%	37.71%	10.14%	25.55%
Range	69	145	31	50
Minimum	13	36	67	29
Maximum	82	181	98	79
1st quartile	32	67	77	46
median	46	82	82	54
3rd quartile	61	108	88	61
interquartile range	29	42	12	15
Mode	32	75	81	50
Kurtosis	-1.03	0.70	-0.62	-0.53
Skewness	0.068	0.684	0.076	0.035
zmin=	ME= -1.64	MK and RS = -1.56	BG= -1.86	RO= -1.80
zmax=	UK= 1.64	IE= 2.86	NL= 1.84	NL= 1.85

Source: Eurostat, Authors' creation

For EU-28 countries, an increase of the main variable studied, with the linear trend slope of 2.73 percentage points yearly over 2007 to 2018, and 99.50% of total variation explained by the model is shown in Figure1. It is visible that the percentage of individuals ordering goods and services increased from 30% in 2017 to 60% in 2018. According to e-commerce statistics for individuals, Eurostat (2019), data for the EU-28 countries, Montenegro, FYR of Macedonia and Serbia, sorted downward by the percentage of those who purchased online, and this was compared to the percentage of those who used the Internet in the same period in 2018. Denmark leads with 98% for those who purchased online within the last 12 months, and 84% for those who used the Internet within the last 12 months with. Montenegro is with 12% at the bottom for those who purchased online within, while for the percentage of those who used the Internet Bulgaria appeared with 67% as being at the bottom.

All four variables, Y2017IntOrderGoods, X1_GDPpcPPS, X2_AccHome and X3_DigitalSkill, were explored and described statistically, Table 1. The highest variability of data over 31 countries is calculated for $Y_{2017IntOrderGoods}$, with Coefficient of variation of CV=44.33%, and the lowest variability was noticed for $X2_{\_AccHome}$ (CV=10.14%). Regardless the high outlier for the variable $X1_{GDPpcPPS}$ for Ireland, with standardized value of Z_{IE}= 2.86, it was remained for the further analysis. The same moderate low extremes were found for FYR of Macedonia and Serbia (Z_{MK}= Z_{RS}= -1.56).

The correlation matrix for 31 countries in 2017, Table 2, shows correlations between all the pairs of variables under consideration, being all positive and strong.

Table 2. *Correlation matrix, for n=31 countries and 2017*

	$Y_{2017IntOrderGoods}$	$X1_{\_GDPpcPPS}$	$X2_{\_AccHome}$	$X3_{\_DigitalSkill}$
$Y2017_{IntOrderGoods}$	1.0000			
X1_GDPpcPPS	0.7654	1.0000		
X2_AccHome	0.9027	0.8129	1.0000	
$X3_{\_DigitalSkill}$	0.9166	0.6852	0.8471	1.0000

Source: Eurostat data, Authors' creation.

The OLS MLR modelling was performed for 31 countries in 2017. Among several models built and tested for validity regarding explanation the dependent variable, $Y_{2017IntOrderGoods}$, the following two multiple regression models were fully developed and interpreted: In MLR **Model 1**, for explanation the dependent variable, $Y_{2017IntOrderGoods}$, two independent variables were used: $X1_{\_GDPpcPPS}$ and $X3_{\_DigitalSkill}$; In MLR **Model 2**, for explanation the dependent variable, $Y_{2017IntOrderGoods}$, two independent variables were used: $X2_{\_AccHome}$ and $X3_{\_DigitalSkill}$. The estimated **Model 1** looks as given in (3):

$$\hat{Y}_{2017IntOrderGoods} = -27.86 + 0.17 \cdot X1_{GDPpcPPS} + 1.14 \cdot X3_{DigitalSkill} \;,$$
$$\hat{\sigma} = 7.683; R = 0.936;\ R^2 = 0.876; \bar{R}^2 = 0.867; F = 98.58; n = 31 \qquad (3)$$

For one index point increase in the variable GDP per capita in PPS, Index, EU28 = 100, $X1_{\_GDPpcPPS}$, with the other independent variable unchanged, the regression value of $Y_{2017IntOrderGoods}$, would increase by 0.17 percentage points. For one percentage point increase in the variable $X3_{\_DigitalSkill}$, having the remaining independent variable fixed, the regression value of $Y_{2017IntOrderGoods}$ would increase by 1.14 percentage points.

The F-Test for overall regression, with p-value=2.12E-13, shows statistical significance of the whole Model 1 at even 1% significance level. Coefficient of determination R^2 tells that 87.6% of the total sum of squares is explained. The individual two-sided t -Tests, shows that variable $X1_{\_GDPpcPPS}$ is statistically significant at 1% significance level (p-value=0.009). The variable $X3_{\_DigitalSkill}$ is statistically significant at 1% significance level (p-value=8.52E-09). No model assumptions violation is present.

The estimated **Model 2** is given in (4):

$$\hat{Y}_{2017IntOrderGoods} = -89.61 + 0.12 \cdot X2_{X2_{AccHome}} + 0.83 \cdot X3_{DigitalSkill} \;,$$
$$\hat{\sigma} = 7.008; R = 0.947;\ R^2 = 0.897; \bar{R}^2 = 0.889; F = 121.30; n = 31. \qquad (4)$$

For one index point increase in the variable $X2_{\_AccHome}$, with the other independent variable unchanged, the regression value of $Y_{2017IntOrderGoods}$, would increase by 0.17 percentage points. For one percentage point increase in the variable $X3_{\_DigitalSkill}$, having the remaining independent variable fixed, the regression value of $Y_{2017IntOrderGoods}$ would increase by 0.83 percentage points. Coefficient of determination R^2 shows that 99.7% of the total sum of squares is explained by Model 2. Since the overall F-Test has p-value = 1.61E-14, the whole regression Model 1 is statistically significant at 1% significance level. Using two-sided t-Test, the variable $X2_{\_AccHome}$ is statistically significant, with t-statistic = 3.908 and p-value = 0.0005, at 1 % significance level. The variable $X3_{\_DigitalSkill}$ is statistically significant, with p-value = 0.0001, at 1% significance level, too. Regression diagnostics' tests for residuals were performed, showing no assumptions violation is present.

Cluster analysis: In the next step, for 31 countries in 2017, based on all four variables examined in the regression modelling, Y2017IntOrderGoods, X1_GDPpcPPS; X2_AccHome, and X3_DigitalSkill, cluster analysis using Ward linkage and Squared Euclidean distances, according to Hair et al. (2008), and Field (2011), was performed, Table3.

Table 3. *Hierarchical clustering with Ward linkage and squared Euclidean distances, 2017.*

Cluster	No. of countries; n = 31	Countries grouped into the clusters
Cluster 1	10	Belgium, France, Austria, Czech R., Slovakia, Estonia, Spain, Malta, Slovenia, Ireland
Cluster 2	6	Denmark, Germany, Finland, United Kingdom, Netherlands, Sweden
Cluster 3	8	Bulgaria, Serbia, Greece, Croatia, Montenegro, Romania, FYR of Macedonia, Turkey
Cluster 4	7	Italy, Cyprus, Portugal, Lithuania, Latvia, Hungary, Poland

4. Discussion and Conclusion

Individuals using the internet for ordering goods or services, as the percentage of individuals aged 16 to 74, $Y_{2017IntOrderGoods}$, which doubled for the EU-28 countries from 30% in 2007 to 60% in 2018, resulted with the highly representative estimated linear trend, with the yearly slope of 2.73%. For the period from 2007 until 2015, such a trend slope was a little bit higher, 2.82%. Because of the highest correlation, r = 0.9166, both here analysed MLR models, built for explanation of the main variable under study, included the Digital Society related indicator named Percentage of individuals aged

1674, with basic or above basic overall digital skills. Additionally, Model 1 included GDP per capita in PPS, while Model 2 included additionally Level of internet access for households, performing the second regressor. Both models, shown that an increase in each of the independent variables would result with a statistically significant increase in the regression value of the main variable under study at 1% significance level. Clustering method resulted with four clusters. Regarding all four variables studied here, as the result of clustering, highly developed countries, Belgium, France, Austria, Czech R., Slovakia, Estonia, Spain, Malta, Slovenia, Ireland, gathered in Cluster 1. The countries, being economically and digitally the most developed ones, Denmark, Germany, Finland, United Kingdom, Netherlands and Sweden, clustered in Cluster 2, separately. These countries are leaders regarding the Digital Society indicators. Eight SEE countries, the EU-28 members plus four EU candidate countries (Bulgaria, Serbia, Greece, Croatia, Montenegro, Romania, FYR of Macedonia, Turkey, with exception of Cyprus), clustered in the same, Cluster 3. Cluster 4 collected Italy, Cyprus, Portugal, Lithuania, Latvia, Hungary and Poland.

More efforts for improved education regarding the prerequisite for better performance of Digital Society over European countries, which includes internet usage and digital literacy, based on improved feeling for financial transactions security, should come from various sides, from governments, educators, employers, employees, citizens, and from national statistical societies. So, governments should offer more adequate legal frame and support for it; educators might should develop and adopt adjusted educational programs; employers might show more understanding by giving more opportunities for additional education on digital and related types of literacy, such as statistical, media and financial literacies; and employees', facing some skill gaps, should pose request for continuous adults' education not only for digital, but also numerical, statistical and financial literacy. Regarding digitisation and globalisation, for what here studied main variable studied (Percentage of individuals who ordered goods or services via Internet), is just an example, the national statistical societies (NSAs) might have an extra challenge and could be of higher influence. NSAs' role at national, European and international level might be of higher importance and more visible, if they could find the way to support "digital skills – based" statistical literacy improving and spreading. Especially those NSAs, which, either research and analyse statistical literacy indicators and respective education process development, trying to influence them, or directly encourage and support up-to-date "digital skills – based" education programs in statistics, might contribute to an increase of Digital Society related indicators (Internet Ordering, Purchases or other similar), through rising voice pro more adjusted programs, not solely regarding statistical, but also digital, media and financial literacies, truly needed for better adoption of Digital Society benefits in modern globalized World per se, as well.

In the future desk research, the limitations of this study may be overcome if more variables, e.g. gender and age groups, as well as different educational levels, should be analysed for better explanation of digital society related indicators, additionally. Further, the impact of the NSA's activities in the European countries may be analysed and measured, by analysing the existence and the efficiency of their attempts to contribute to the improvement of statistical and related digital literacy, through promotion of educational programs, which should nowadays be more often visible or offered at the NSA websites. Recently, since the beginning of 2019, FENStatS (2019) has been announcing (parallel to the announcing the courses organized by European Courses on Advanced Statistics, ECAS), the upcoming NSAs' Statistical and Data Science conferences more often, which is based on the improved overall digital skills adoption. This activity is encouraging and supportive to the NSAs' acting goals and efforts in the same direction to reach better informed, "digital based" statistical literacy improved, for better citizenship and society.

Here recognized clusters of countries, with significantly different development levels and percentages of individuals for $Y_{2017IntOrderGoods}$, and $X3_{\_DigitalSkill}$, which have been taken in this research as the symbols for digitalisation and globalisation, have also more or less active communication habits of their NSAs. Because of the research time constrains, their activity and communication style were not explored with full potential for this paper, but generally speaking, each NSA's activity alone and towards the FENStatS could be in many countries intensified through regular standardized communication procedure and praxis, being regularly current (e.g. on quarterly or similar basis), efficient and formalized. Proposal for a common society, such as FENStatS, is directed to an idea of intensifying communication between NSAs by using the existing FENStatS' web site, based on standardized information collection and current information refreshment, which would enable information announcement, flow and exchange. Since such an approach should be simple, a short standardized format could be suggested for use. Current news of interest might include the categories: Statistical and Data Science education and training programs; including statistical and digital knowledge and literacy improvement possibilities; Conferences organised by NSAs Call for Papers announcements; and Other news. This would improve networking among the CSAs' members and the overall importance of the FENStatS' activity, based on practical use of available ICT with an increasing potential for improvements in literacy and knowledge regarding Statistics and Data Science in the World of digitalisation and globalisation.

References

1. CB (2017). Consumer Barometer with Google https://www.consumerbarometer.com/en/about/ (Accessed 18 March 2019)
2. Dumičić, K., Čeh Časni, A. & Žmuk, B. (2015). Internet Purchases by Individuals: Clusters of the Most vs the Least Developed European Countries. *Int. J. Inf. Tech. Decis.*, **18**, 1–9.
3. Dumičić, K., Skoko Bonić, I. & Žmuk, B. (2018). Statistical analysis of the development indicators' impacts on e-commerce of individuals in selected European countries. *Our economy*, **64**(2), 15–24.
4. Dumičić, K., Žmuk, B. & Čeh Časni, A. (2015). Impacts on travel and holiday accommodation online booking in selected South-East European countries: a multivariate analysis approach. *Croat. Oper. Res. Rev.,* **6**(1), 155–168.
5. Dumičić, K., Žmuk, B. & Mihajlović, I. (2017). Profile analysis of clustered European countries based on selected variables impacting the e-commerce realized by individuals. In *ISI World Statistics Congress-WSC, ISI2017. In press.*
6. Dumičić, K., Žmuk, B. & Novkovska, B. (2017). Regression analysis of e-commerce: focusing the selected EU candidates and the EU countries. Knowledge and business challenge of globalisation in 2017: conference proceedings of the 7th international scientific conference. Celje, Slovenia, November 17, 2017. FKPV, 156–164, CD. Electronic ISBN: 978-961-6997-04-1.
7. Dumičić, K., Čeh Časni, A. & Palić, I. (2015). Multivariate analysis of determinants of Internet banking use in European Union countries. *Cent. Eur. J. Oper. Res.*, **23**(3), 563–578.
8. Eurostat (2019). https://ec.europa.eu/eurostat/ (Accessed 20 May 2019) FENStatS (2019). FENStatS News. http://www.fenstats.eu/news (Accessed 20 May 2019).
9. Field, A. (2011*). Discovering statistics using SPSS*, 3rd Ed. London: Sage.
10. Gujaraty D. N. & Porter D.C. (2010). Essentials of Econometrics 4th Ed., Irwin: McGraw-Hill.
11. Hair F.H., Black W.C., Babin B.J., & Anderson R.E. (2008). *Multivariate Data Analysis*. 7th Ed., Upper Saddle River, NJ: Prentice Hall.
12. Nagy, S. (2016). E-commerce in Hungary: A Market Analysis. *Club of Economics in Miskolc TMP*, **12**(2), 25-32.
13. Nagy, S. (2017). Digital economy and society–a cross country comparison of Hungary and Ukraine. *arXiv preprint arXiv:1901.00283.* Žmuk, B., Dumičić, K. & Mihajlović, I. (2014). On–line booking use for travel and holiday accommodation and development indicators: Clustering of European countries. *Proceedings of 7th International*

Scientific Conference "Economic and Social Development" –ESD, New York, USA, October 24, 2014, 508–518.
14. Žmuk, B. & Mihajlović, I. (2018). Online booking for travel and accommodation influenced by economic and digital development level: Position of the Western Balkan countries within Europe. *Croatian Review of Economic, Business and Social Statistics (CREBSS)*, **4**(2), 86–98.

The new proposal of forest products classification and definitions

Arvydas Lebedys[1], Michal Synek[2]
[1] Forestry Officer (Statistics), FAO, Rome
[2] Member of the national team on Forest Products Statistics in the Czech Republic, Olomouc – Czech Republic

Abstract

The first version of Forest Products Classification and Definitions (FPC&D) was prepared by Food and Agriculture Organization of the United Nations (FAO) and published in 1973 with last adoption in 1981. The paper summarizes the updated proposal on classification and definitions that takes into account the recent changes in technology, changes in wood-based industry and in trade and incorporate new wood and paper products.

The new proposal of FPC&D constitutes a complete product classification covering all wood and paper products. It has been designed to provide a comprehensive view on wood and paper products which will be useful in relation to forestry and forest industry production, industry capacity and trade. They do not attempt to lay down strict definitions such as might be found in national standards or in trade and tariff regulations, but they do aim at drawing attention to the characteristics which differentiate between items that are of practical importance in the different areas of the sector's activities, and to indicate aggregate and subdivisions of products and materials that have practical significance. The new version of FPC&D also provides a framework for international comparison and promotes harmonization of various types of statistics related to forest products. Moreover it can be useful in developing statistical systems related to forests and producing statistics on production, industry capacity and trade and it can also serve as a useful reference for the correspondences between existing international statistical systems related to forests.

Keywords

forest products statistics; wood products; paper products

1. Introduction

A system of classification and definitions is an essential component in work on forest products statistics, industry and trade statistics, and in the development of forestry and industry statistical systems. A variety of classifications, nationally and internationally, have been developed to suit a wide range of statistical needs. Therefore, there is also a strong need for correspondence tables between various classification systems. Meanwhile, these classifications are subject to revisions and changes over time because a statistical classification must be up-to-date as recommended by United Nations Statistical Division (UNSD) (Hoffmann & Chamie, 2002).

The first version of the FPC&D was published in 1973. This version was superseded by the second version, published in 1979 and the last version was adopted in 1981 (FAO & UNECE, 1981) and it is valid until now. A unique numerical reference to each element and aggregate was used to identify characteristics used for classification of subdivisions, namely stage of manufacture, chief component material, technology applied and intended use. Regarding coverage, the FPC&D cover the wood and wood-based products for which FAO and United Nations Economic Commissions for Europe (UNECE) collected statistics on a regular basis in 1981. Very important and useful part of the FPC&D is the harmonization with the major international classification by means of cross references, which are defined as indications of some degree of coincidence between the two classifications, not of a one-toone correspondence.

Other major international classifications, used mainly in trade statistics, describe wood and paper products in more or less detailed level.

Harmonized System (HS) was developed by World Customs Organization (WCO) as an international nomenclature for the classification of products. According to Ramaschiello (2015) "more than 98% of the merchandise in the international trade is classified using HS and almost all countries use it for statistical and administrative purposes". It is revised every five years and next revision HS 2022 will be released in 2022. The WCO sets a threshold of USD 50 million of annual trade value for new subheadings and USD 100 million for new headings (Ramaschiello, 2015). The classification in the HS is based on the next criteria: (a) raw or basic material, (b) degree of processing, (c) use or function, (d) economic activity. For the followed HS revision (HS 2022), next new 50 subheadings are proposed by the Intersecretariat Working Group (IWG) in order to improve classification of wood and paper products in the international trade.

The Central Product Classification (CPC) covers all goods and services, including transportable goods, non-transportable goods and services. Distinguishing characteristics are the physical properties and the intrinsic nature of the products, these include, for example, the raw materials of which goods are made, the stage of production or the way in which goods are produced or services rendered, the purpose or user category for which products are intended and the prices at which they are sold. The industrial origin is distinguished on the level of subclasses (UN, 2015). The CPC is revised every five years. The latest version CPC Ver.2.1, which has been released in 2015, encompasses 21 new five-digit codes for forest products and it is better harmonized with HS 2012 compared to previous one.

Standard International Trade Classification (SITC), which last revision SITC Rev.4 was officially released in 2008 covers all goods classifiable in HS 2007, except for monetary gold, gold coin and current coin. SITC is now recommended only for analytical purposes, so there has been no need to revise the structure according to valid HS. The commodity groupings of SITC reflect: (a) the materials used in production, (b) the processing stage, (c) market practices and uses of the products, (d) the importance of the commodities in terms of world trade, and (e) the technological changes (UN, 2015). International Standard Industrial Classification (ISIC), which last revision ISIC Rev.4 (UN, 2008) has been released in 2008, is mainly focused on the industrial origin. It means on: (a) the inputs of goods, services and factors of production; (b) the process and technology of production; (c) the characteristics of outputs and the use to which the outputs are placed. It is a classification according to the kind of productive activity, and not a classification of goods and services, for which cross-references to CPC are defined.

Broad Economic Categories (BEC) classification was originally designed to be used for the compilation of data on international trade by large economic classes of commodities. It was also designed to serve as a means of converting external trade data compiled on SITC into end-use categories that are meaningful within the framework of the System of National Accounts (SNA), namely, categories approximating the three basic classes of goods in SNA: capital goods, intermediate goods and consumption goods. (UN, 2011).

On the national level, classification of wood and paper products are rare, but as an example Canada can be mentioned, where A New Taxonomy of Wood Products (Cohen, 1996) and A Revised Taxonomy of Wood Products (Cohen, 2008) were published. The first taxonomy covered both composite products and solid wood but does not take into account products of chemical processing (pulp and paper). The later taxonomy categorizes both composite products and solid wood but quickly summarizes also pulp and paper products.

2. Methodology

In February 2015 at its 29th session, the Intersecretariat Working Group on Forest Sector Statistics (IWG), which is composed of four active members EUROSTAT, FAO (Forestry Department), International Tropical Timber Organization (ITTO) and UNECE (Forestry and Timber Sector) recommended to update the Classifications and Definitions of Forest Products publication to accommodate recent changes in the major international classifications and forest products markets.

Recommendation included to maintain a similar structure as the previous versions, while incorporating recent changes to the Joint Forest Section

Questionnaire (JFSQ) definitions and classification, and to refer to the 7th edition of the General Nomenclature of Tropical Wood (ATIBT, 2016). The updated classification was expected to take into account not only products included in the JFSQ but also other wood and paper products that had not been covered by the JFSQ yet (e.g. railway sleepers, densified wood, cork, etc.).

In the first stage, the revision of the FPC&D (FAO & UNECE, 1981) was done in order to evaluate outdated wood and paper products included in the classification and missing new products. HS 2012 and HS 2017, producers' and associations' information materials as well as technical publications (e.g. Forest Products Annual Market Revue) were used as source of information on new wood and paper products. Based on this revision, it was decided to propose a completely new structure and coding system in comparison with the current FPC&D (FAO & UNECE, 1981)

3. Results

The classification is designed to cover the whole spectrum of primary and secondary wood and paper products. It encompasses not only more commonly produced and traded wood and paper products for which FAO collects statistics on a regular basis, but it also covers new wood-based products which appeared during the last years. Included is wood taken from forests or from trees outside the forest, bark and cork; charcoal; wood and wood-based materials resulting from the first processing of the wood available from forest operations (including sawnwood, railway sleepers, veneer sheets, wood pulp and wood residues); and materials resulting from further processing of some of these materials (e.g. woodbased panels, paper and paperboard); recovered paper and recoverable wood products. In line with JFSQ structure, the previous classification was also extended for secondary wood and paper products for which trade statistics are regularly collected.

The classification does not cover manufactured articles of wood and paper and also does not attempt to cover non-wood forest products such as small ornamental trees cut for special purposes; nuts, berries, seeds, or other parts of plants gathered in forest areas; gums, balsams, lacs, etc.; wood derivatives such as turpentine, tall oil, sulphite dye and other chemicals, for which a separate publication was prepared and published by FAO (Sorrenti, 2017). Forestry operations and services are also not taken into account. Non-wood materials are partially included in the new classification, namely in the group panels where other ligno-cellulosic materials are taken into account and in the group pulp which contains pulp made from non-wood fibrous vegetable materials as well as waste paper. The reason for inclusion of these products was, that they have the same use as same material made from wood and may be produced in conjunction with, wood or wood-based materials.

Ten main groups, named Sections, are defined in order to cover each principal stage of operation from the initial felling of the tree to manufacture of the primary product suitable for consumption or further conversion. It also takes into account the fact that the by-products from one stage of operation may be used as the raw material for another.

The Sections have the following distinction:

01 Wood in the rough (Roundwood)

It covers all wood raw material taken from wood in the rough, including chips produced directly in the forest. The basic division adopted for the classification in this section is on wood fuel and wood in the rough other than wood fuel.

02 Wood simply worked or processed

This section includes wood charcoal, torrefied wood, roundwood treated with preservatives, roughly trimmed wood suitable for the next manufacture or direct use, wood wool, wood flour, wood fuel simply worked and other wood simply worked such as staves, shingles and shakes.

03 Wood chips and particles, residues and recoverable wood products

Wood chips and particles covers intermediate products, which may be manufactured from a number of sources (wood in the rough, residues or recovered wood products) and have a great variety of uses e.g. for pulp, particle board, etc.

Wood residues consist of wood which has passed through some form of processing but which also constitutes the raw material of a further process e.g. for pulp, particle board, etc.

Recoverable wood products are incorporated into classification as potential source of the raw material for production of pulp, particle boards, fibreboards or for energy purposes.

04 Wood pellets and other agglomerates

This section includes wood pellets and other agglomerates such as wood briquettes which are mainly produced from co-products of the mechanical wood processing industry, furnituremaking industry or other wood transformation activities.

05 Sawnwood

This section covers the products of the simple processes of sawing lengthways, with the associated processes of hewing or profile chipping. It encompasses sleepers, sawnwood rough or sawnwood planed

06 Veneer sheets

Veneer sheets are thin sheets of wood of uniform thickness not exceeding 6 mm obtained by sawing, slicing or peeling. Veneer sheets are used for manufacture of plywood, laminated construction materials, veneering furniture, etc.

07 Wood-based panels (including panels from other ligno-cellulosic materials)

Wood-based panels covers different products, such as plywood, particle board, oriented strandboard (OSB), fibreboard, densified wood, combination board and other panels based on other ligno-cellulosic materials.

08 Wood pulp

Woodpulp is produced from pulpwood, wood chips or residues and it encompasses mechanical pulp, semi-chemical pulp, chemical pulp and dissolving pulp.

09 Other pulp

This section covers the products of processing of the other ligno-cellulosic raw materials (bamboo, straw, bagasse, etc.) and recovered fibre pulp which is manufactured from recovered paper or paperboard.

10 Recovered paper

This section covers waste and scraps of paper or paperboard that have been collected for reuse or trade.

11 Paper and paperboard

Graphic papers, sanitary and household papers, packaging materials and other paper and paperboard are included in this Section.

12 Cork

This section incorporates raw, semi-processed and worked cork.

13 Secondary wood products

Further processed sawnwood, wooden wrapping and packaging material, wood products for domestic or decorative use, other manufactured wood products, builder's joinery and carpentry of wood, wooden furniture and prefabricated buildings are classified under this Section.

14 Secondary paper products

This Section covers further processed paper and pulp, namely composite paper and paperboard, special coated paper and pulp products, carbon paper and copying paper ready for use, household and sanitary paper ready for use, packaging cartons, boxes and other articles of paper and paperboard ready for use

The classification structure is based on the followed criteria:

a) Wood origin for the primary wood products
b) Wood species origin
c) Stage of manufacture

d) Chief component material
e) Technology applied in the production process
f) Purpose or intended use

The variability of wood, paper and paperboard products does not allow to use the same criterion on the same level in each Section of forest products. The classification criteria were then structured according to their importance for classification of particular Section of forest products. A new six-levels coding system has been proposed to take into account all criteria.

4. Discussion and Conclusion

The new proposal of Forest Products Classification and Definitions (GSARS, 2016) has been published as a working paper in order to serve as a living document, which is open for all relevant comments, suggestion and proposals.

It is designed to provide a comprehensive view on wood and paper products which will be useful in relation to forestry and forest industry production, industry capacity and trade. It does not attempt to lay down strict definitions such as might be found in national standards or in trade and tariff regulations, but it does aim to draw attention to the characteristics which differentiate between items that are of practical importance in the different areas of the sector's activities, and to indicate aggregate and subdivisions of products and materials that have practical significance.

The new classification, in accordance with the previous classification (FAO & UNECE, 1981), takes into account all parts of the tree, because each part of the tree is potential raw material for forest products. Moreover, it takes into account secondary wood and paper products. Regarding wood species origin, the previous classification, compared the new one, does not distinguish non-coniferous tropical wood origin. The new classification divides non-coniferous wood species into two items (tropical nonconiferous, other non-coniferous) by all sections and divisions where it is important for their classification (e.g. 06 Veneer sheets).

Very important and useful part of the new classification, as well as the previous one, is the harmonization with the major international classification by means of cross references, which are defined as indications of some degree of coincidence between the two classifications, not of a one-toone correspondence. The selected international classification systems are HS 2012, HS 2017, SITC Rev.4, CPC Ver.2.1, ISIC Rev.4. Moreover, correspondences with Joint Forest Section Questionnaire (JFSQ), which is used by IWG as a main tool for forest products statistics data collection on the international level, were added. As a results, the new classification in connection with Guidelines on data collection for national statistics on forest products (GSARS, 2018), can provide developing countries with the technical guidance how to develop and/or improve their national forest products statistical programmes.

In comparison to HS, the new classification provides more detailed classification of primary and secondary wood products. Regarding to the primary and secondary paper products the classification more or less follows HS structure.

Also in comparison to CPC Ver.2.1, the new classification provides more detailed classification of wood and paper products. CPC Ver.2.1 is more oriented to industrial origin and takes into account also services; so it is more suitable for System of National Accounts (SNA) and international comparisons.

References

1. Cohen, D.H., Ellis, S.C., Kozak, R.A., Wilson, B. (1996). A New Taxonomy of Wood Products. Pacific Forestry Centre Canadian Forest Service, Available at: http://cfs.nrcan.gc.ca/pubwarehouse/pdfs/4351.pdf Accessed February 2019.
2. Cohen, D.H., Ellis, S.C. (2008). A Revised Taxonomy of Wood Products. Pacific Forestry Centre Canadian Forest Service
3. Hoffman, E., M. Chamie. (2002). Standard statistical classifications: basic principles. Statistical Journal of the United Nations Economic Commission for Europe Vol.19(4):223241. FAO: Rome,
4. Ramaschiello, V. (2015). FAO Contribution to the Harmonized System 2017. New York. USA. Available at: https://unstats.un.org/unsd/classifications/expertgroup/egm2015/ac289-15.PDF Accessed July 2016
5. Association Technique Internationale des Bois Tropicaux (ATIBT), (2016). General nomencalture of tropical timber, 7th edition., Association Technique Internationale des Bois Tropicaux, Nogent-sur-Marne Cedex, Available at: http://www.fao.org/3/a-be999f.pdf Accessed February 2019
6. Food and Agriculture Organization of the United Nations (FAO), Global Strategy for Improving Agriculture and Rural Statistics (Global Strategy) and United Nations (UN), (2015). Guidelines on International Classifications for Agriculture Statistics. Available at: http://gsars.org/wp-content/uploads/2015/12/Guidelines-for-Int-Classifications-onAgricultural-Statistics-web.pdf Accessed February 2019.
7. Food and Agriculture Organization of the United Nations (FAO), United Nations Economic Commission for Europe (UNECE), (1982). Classification and definitions of forest products. Available at: http://www.fao.org/3/a-ap410m.pdf Accessed February 2019.

8. Sorrenti, S. (2017). Non-wood Forest Products in International Statistical Systems. Available at: http://www.fao.org/3/a-i6731e.pdf Accessed February 2019.
9. The Global Strategy for improving Agriculture and Rural Statistics (GSARS). 2018. Guidelines on data collection for national statistics on forest products. Available at: http://gsars.org/wp-content/uploads/2018/12/GS-NFP-GUIDELINES-EN-06.pdf . Accessed February 2019.
10. The Global Strategy for improving Agriculture and Rural Statistics (GSARS). 2016. Forest Products Classification and Definitions. Working Paper No. 14. Available at: http://gsars.org/wp-content/uploads/2016/12/WP-23.12.2016-Forest-Products-Classificationand-Definitions-MSALv4.pdf . Accessed December 2017.
11. United Nations. (2002a). Standard International Trade Classification. Available at: http://unstats.un.org/unsd/iiss/Print.aspx?Page=Standard-International-Trade-Classification. Accessed July 2016.
12. United Nations. (2002b) International Standard Industrial Classification of All Economic Activities (ISIC), Revision 3.1. Available at:http://unstats.un.org/unsd/publication/seriesm/seriesm_4rev3_1e.pdf. Accessed February 2019.
13. United Nations. (2006) Standard International Trade Classification, Revision 4.(2006). Available at: http://unstats.un.org/unsd/publication/SeriesM/SeriesM_34rev4e.pdf Accessed February 2019.
14. United Nations. (2008) International Standard Industrial Classification of All Economic Activities (ISIC), Revision 4. (2008). Available at: http://unstats.un.org/unsd/publication/seriesM/seriesm_4rev4e.pdf Accessed July 2016.
15. United Nations. (2011) Revision of the Classification by Broad Economic Categories (BEC).(2011). Available at: https://unstats.un.org/unsd/publication/SeriesM/SeriesM_53rev4e.pdf Accessed July 2016.
16. United Nations. (2015) Central Product Classification (CPC), Version 2.1.(2015). Available at: https://unstats.un.org/unsd/classifications/unsdclassifications/cpcv21.pdf Accessed February 2019.

Strategic Planning for Agriculture and Rural Statistics (SPARS) and its impact at country level

Arbab Asfandiyar Khan[1], Hem Raj Regmi[2]
[1] United Nations Food and Agriculture Organization, Rome, Italy
[2] CBS Nepal and International Consultant at FAO, Kathmandu, Nepal

Abstract

Evidence based planning, monitoring and decision making are the key for the success of national development process. It is more evident in those countries where majority of the population are engaged in agriculture and live in rural areas. Considering the low profile of the agriculture statistics in many developing countries of Africa, Asia and Americas; UNFAO in collaboration with UNSD, UNSIAP, ADB, AfDB and many other development partners implemented the project named Global Strategy to Improve Agricultural and Rural Statistics (GSARS) in more than 70 countries of the world and about 19 countries in Asia Pacific region.

One of the basic but most important outcome of the Global Strategy project is the Strategic Plans for Agricultural and Rural Statistics (SPARS), which is based on the In depth country assessments (IdCA) of the project implementing countries and focuses on the immediate to long term activities for the proper development of agriculture statistical systems. SPARS include agriculture statistical system in broader sense (i.e. crop, livestock, fisheries and forestry sectors) of the agriculture statistics and identifies the vision, mission, goal, objectives and major prioritized and actionable activities to overhaul the system and better coordinate it across the National Statistical System. It also focuses on the financial strategy with proper monitoring and evaluation plan for its implementation.

Some significant impact has been observed during the implementation phase of the GSARS in the Asia Pacific region. Few examples include the establishment of Ministry of Statistics in Pakistan, establishment and expansion of Agriculture statistics division at ministry of agriculture and forestry in Bhutan, joint (ministry and NSO) crop production estimation system established in Bangladesh, Sri Lanka and Indonesia amongst others. This paper illustrates the major impacts achieved by the SPARS implementing countries based on the country reports and communication material i.e. newsletters published by GSARS project..

Keywords

Agricultural statistics; Statistical Plan; Agricultural Statistics strategy

1. Background

Transition from the Millennium Development Goals (MDGs) to the Sustainable Development Goals (SDGs) has obligated sound national statistical systems to provide the data needed to monitor progress towards meeting these goals. The quality and availability of agricultural statistics is of paramount importance in measuring the progress to achieve the SDGs related to food and agriculture. However, in the absence of any long term national strategy for the development of statistics in general and/or agricultural statistics in particular, necessitate countries to improve their agricultural statistics system for the production of adequate quantity and better quality agricultural statistics.

The lack of vision and political will for improving statistics in many developing countries steered the strategic planning process in the statistics production. While the Partnership in Statistics for Development in the 21st Century (PARIS21) advocated the National Strategy for the Development of Statistics (NSDS) for the overall development of the national statistical systems, the Global Strategy to Improve Agricultural and Rural Statistics (GSARS) established a comprehensive framework for improving the quality, availability and use of agricultural and rural data.

The Global Strategy as an international partnership was funded by Gates foundation, DFID AND Italian cooperation and was administratively housed by the Economic and Statistics Division of Food and Agriculture Organization of the United Nations (FAO). The ultimate aim of the project in its phase 1 was to improve the steady decline in the availability and quality of agricultural and rural statistics in developing countries. The Global Strategy provides a framework to enable national and international statistical systems to produce the basic agricultural information needed to guide decision making.

The Strategy is based on three pillars:

Pillar 1: Establish a minimum set of core data required to meet current and emerging needs.

Pillar 2: Integrate agriculture into national statistical systems and use sound data management systems; and

Pillar 3: Establish suitable governance processes and build the necessary statistical capacity to ensure sustainability of agricultural statistics systems.

This paper highlights the major technical areas outlined in SPARS and its progress in 19 implementing countries of Asia Pacific region from 2013 to 2019. (Note: The SPARS guidelines and cost effective methodologies and technical areas highlighted below were also delivered to countries that were not covered by GSARS).

2. Strategic Planning for Agriculture and Rural Statistics (SPARS)

SPARS is a Long Term Strategy to improve Agriculture and Rural Statistics (ARS) at the national level which derives its basis from National Strategy for the Development of Statistics (NSDS). This is the building block of NSDS and can be linked to development Planning process of the country with the following objectives

- Understanding the impact of agriculture policy on statistical priorities
- Identifying data needs, gaps, deficiencies, duplications and inconsistencies
- Defining future short- and long-term statistical programs and interventions

SPARS design stages: The SPARS design process includes following three phases

1. **Launching/Preparing Phase:** which includes the Understanding, acknowledging, committing, Building constituency, Integrating the SPARS in the NSDS and Roadmap
2. **Assessment Phase:** Detail assessment of the key sectors of Agriculture and Rural Statistics is conducted and IdCA report is prepared during this phase.
3. **Planning Phase:** This phase includes the development of Vision and Mission, Strategic Goals and Outputs and Action Plan with financial and Monitoring frameworks.

Priority areas identified as a result of the SPARS process:

The major technical areas covered in SPARS and their status during implementation of the project are highlighted below.

1. **Institutional Development and coordination mechanism**

One of the most important objective of the SPARS was either to establish an institutional setup, where it was non-existent or develop the capacity of the institutions if they already existed. Following are the major milestones achieved on this front during the SPARS implementation.

- Active high level steering committees (led by secretary or director general), established in almost all project implementing countries. The committees has as representatives the major producers and users of agriculture statistics at the country level.
- Functional Technical Working groups established in almost all project implementing countries which includes the members of major producers of agriculture statistics including Federal and Provincial Government i.e. NSOs, MOAs, other ministries and the data users i.e. Academia, Research Institutions and civil society s.
- Establishment of Ministry of Statistics in Pakistan

- Establishment of the Agriculture Statistics Division in Ministry of Agriculture and Forests in Bhutan
- Expansion of the agriculture division in Bangladesh
- Joint committees established for the estimation of crops in Bangladesh, Indonesia, Myanmar and Sri Lanka

2. Improving the quality of Administrative Reporting Systems

In some countries where there were no established census or survey systems, administrative reporting systems are the key source of statistics including agriculture statistics. Major activities performed in this technical area and progress achieved are,

- Improving Administrative Reporting Systems for Rice Production Statistics in LAO PDR and Cambodia.
- Awareness raising through workshops in all implementing countries to improve the quality of the administrative data to make them statistically reliable and valid.

3. Computer-assisted personal interviewing (CAPI)

CAPI is an interviewing technique in which an electronic device is used to carry out the survey. It is similar to telephone interviewing, except that the interview takes place in person instead of over the telephone. This method is usually preferred over a telephone interview when the questionnaire is long and complex. It has been classified as a personal interviewing technique because an interviewer is usually present to serve as a host and to guide the respondent. CAPI is cost effective as it saves the costs of printing questionnaires, data entry, editing, coding and shortens the overall time for the data collection

Global Strategy in collaboration with World Bank has developed the software called Survey Solutions to assist in implementing the CAPI in beneficiaries' countries.

- Bhutan Agriculture Census,
- Mongolia in livestock survey,
- Timor Leste, Agriculture Census,
- Maldives, Male market data among others
- Viet Nam in agriculture survey

4. Cost of Production

The Cost of Production (COP) estimates are used to guide and inform farmers and agriculture policy makers in making investment decisions and policies. CoP manual was developed and implemented in various countries focused as follows

- Cost of Production of Mango in Bangladesh
- Cost of Production of Rice in Indonesia and Viet Nam

- Cost of Production of Milk and meat in Mongolia

5. Master Sampling Fram e

Master Sampling Frame provides the basis for all data collections through sample surveys and censuses, by allowing for selection of samples for several different surveys or different rounds of the same survey, as opposed to building an ad-hoc sampling frame for each survey.

Thanks to the design of SPARS, Bangladesh, Maldives and Timor-Leste have already developed their MSF and others countries like Sri Lanka and Nepal are in the process of development.

6. Food Balance Sheet

Food Balance Sheets present a comprehensive picture of the patterns of a country's food supply during a specified reference period. The food balance sheet shows for each food item the sources of supply and its utilization. Trends in the overall national food supply can be identified through regular tabulation of annual food balance sheets. These tabulations can also identify changes that may have taken place in the types of food consumed, i.e. diet patterns, and reveal the extent to which the food supply of the country as a whole is adequate in relation to nutritional requirements. By combining food and agricultural data in a country, a food balance sheet can provide a detailed examination and appraisal of its food and agricultural situation. Following 12 countries have drafted/finalized their Food Balance Sheets for one to five years.

- Bhutan, Cambodia, China, Indonesia, Lao PDR, Maldives, Mongolia, Pakistan, Sri Lanka, Myanmar, Nepal and Viet Nam

7. Crop Production Estimation

Among the Minimum Set of Core Data recommended by SPARS, production of data on various crops is the priority activity for many countries particularly those that are dependent on agriculture for the employment and contribution to their economic value added.

Ongoing activities on crop production estimation in 3 countries

- Bhutan (rice yield estimation),
- Fiji,
- Myanmar (rice yield estimation) and

3 other countries have already completed the implementation of crop estimation related surveys.

- Bangladesh (Crop area estimation and yield forecasting [AMIS])
- Indonesia (estimation of rice production [Statistics Indonesia])
- Sri Lanka (estimation of rice production and highland crops)

8. Compilation of Livestock Statistics

This includes the various types of animals and poultry as well as their products like milk, meat and eggs. Through SPARS countries have prioritized these activities, which are now implementation phase in following countries.

- Ongoing activity in Myanmar - Development of the methodology for an Integrated Sample Survey for livestock products
- Completed in Bhutan - Pilot livestock survey on estimation of Milk Production in Haa (2015) province.

9. Pacific Strategic Plan for Agricultural and Fisheries Statistics (PSPAFS)

PSPAFS, is the regional SPARS Jointly launched and implemented by FAO and Pacific Community (Secretariat of Pacific Community, SPC) at the 27th Session of the Asia Pacific Commission on Agricultural Statistics (ACPAS) in March 2018 in Fiji, Nadi. It includes almost all components of the SPARS at regional level focusing on Pacific Island Countries.

10. Others

Regional-level workshops were organized for dissemination of guidelines on: Master Sampling Frame, Post-Harvest Losses, Sub-sector statistics (crops, livestock and fisheries) and Regional trainings on use of ICT (satellite imagery and drones) for agricultural statistics One in Asia, and one in the Pacific

11. Selected Country Status (Success Stories)

To provide additional evidences on the key activities undertaken by country as part of the SPARS process:

I. Afghanistan

- Statistical systems dysfunctional for many years due to civil conflict, revived after the new constitution.
- Current security situation has adverse impact on capacity to produce reliable data.
- Core infrastructure for agricultural statistics in the Ministry through international partnership
- Estimates for wheat, rice, maize and barley by Statistics Department for 21 out of the 34 provinces through the Agriculture Prospects Report.
- Recently satellite data of cropped area for wheat with ground trothing data from field locations used to update area of crops.

II. Bangladesh

- Competencies of NSS strengthened,
- National Quality Assurance Framework for ARS developed
- Core-set of indicators available according to GSARS

- Updated methodologies/best practices followed by other countries adopted
- Improved coordination between BBS and other data producing agencies of ARS
- Harmonized concepts and definitions used by all ministries/agencies of ARS
- More administrative data used as official statistics
- Increased participation of stakeholders for assessing data demands
- Annual survey on input cost of crops, fisheries and livestock conducted using improved methodology
- Annual food balance sheet prepared
- Survey on private food grain stock (supply and value chains) conducted
- Survey on irrigation system conducted
- Survey on land use conducted
- Survey on forest product conducted
- New databases to measure Natural Resources (land, Water and Air) changes developed
- Statistical cell in planning wing of different ministries/departments established
- Training facility created with faculty of agriculture statistics
- Training of personnel involved in Agriculture Statistics provided
- Agriculture wing, BBS reorganized
- Price data system of Department of Agriculture Marketing (DAM) strengthened
- Spatial Data Analysis Unit in Agriculture Wing of BBS created
- Infrastructure for using GIS and Remote sensing techniques developed
- Meta data on each Agriculture and Rural Statistics ARS product available
- Visibility of AR statistical product in public domain improved
- Data sharing protocols developed
- Integrated data dissemination platform established

III. Bhutan

- Pilot milk estimation survey completed in 2 Dzongkhags (like districts)
- Short-term trainings availed for Survey Solutions tool (4 staff), Cost of Production studies and advocacy in agricultural statistics (1

staff), Sampling techniques (2 staff), Food balance sheet (1 staff soon to avail)

- RNR Statistics Division established with 9 staff on 1st August 2017
- RNR Statistics Technical Working Group established in September 2017
- Manual on crop-cutting experiment adopted based on study in 3 Dzongkhags
- Technically sound RNR Census planned to be conducted in Dec 2018

IV. Lao PDR

- Agricultural statistics are based on administrative reporting; data being transmitted weekly from villages to districts to provinces to Ministry of Agriculture and Forests (MAF) central office.
- The agricultural statistics system has the weaknesses inherent in any administrative reporting system
- The administrative reporting system only provides data on crop area/production and livestock numbers.
- There were serious data gaps, such as cost of production, agricultural prices and food stocks.

V. Myanmar

- Food Balance Sheets for the year 2011 to 2016 has been prepared
- Livestock Census was last conducted in 1980
- Subsequently the collection of livestock data on a census basis became a part of the Agricultural Census.
- Though the data from an exclusive livestock census help to better understand the potentialities of the sector to improve the economy, in view of the resources involved, the livestock data will continue to be collected as a part of the agricultural census.
- Crop Production estimation (rice)

VI. Papua New Guanae

- Papua New Guinea Cost of Production (Plantation crops)
- Use of CAPI in CoP and other surveys

VII. Sri Lanka

- Inconsistencies in statistics produced by DCS and the line agencies.
- There are also data gaps, especially crop condition data;
- Joint monitoring system established
- food security indicators through the preparation of food balance sheets and PoU estimated by using HIES data

VIII. Viet Nam

- Statistical Law 2003 updated in 2016.
- Development Plan of the Viet Nam Statistical Development Strategy 2011 to 2020 and Vision to 2025 (VSDS11-20) drafted and under implementation
- Statistical indicator system, updated 2016.
- National statistical survey program, updated 2016;
- Statistical information dissemination policy developed (2016)
- Design Master Sample Frame.

3. Conclusion

Capacity Development: SPARS implementation process has involved key stakeholders from national institutions working on Agriculture and Rural Statistics. Most of them have participated on the regional and global conferences on different thematic areas. Country level workshops and trainings were conducted in all countries to capacitate the people working in ARS sector.

Funding Opportunities: In most of the countries the funding of activities highlighted in SPARS was supposed to be done by the local government. But due to limited resources available locally, many development partners and technical agencies have started funding the SPARS activities in many countries of the region. Few examples are

- Large scale crop production survey in Sri Lanka funded by ADB
- Livestock Survey in Myanmar funded by World Bank
- Joint crop monitoring in Bangladesh funded by ADB
- AGRIS tool is implemented in Cambodia, Nepal and being funded by different partners and coordinated by FAO
- Pre Agriculture Census activities in Bhutan, Bangladesh, Laos, Timor Leste being funded by FAO through TCP
- Major surveys in Agriculture are funded by local governments in Indonesia, Malaysia, China and Pakistan.
- Livestock Census in Mongolia funded by local Government ets.

Challenges encountered during SPARS implementation:

Following are the main observed challenges during the SPARS implementation process in the region

- Uncertain and inadequate financial resources;
- Limitedly skilled and demotivated human resources for developing agricultural statistics;
- Weak institutional and organizational processes;
- Weak and underdeveloped methodological processes;

- Inadequate IT and statistical infrastructure for collecting and analysing agricultural statistics;
- Inappropriate alignment of the statistical offer to demand (national policies and international requirements);
- Lack of adequate capacities to analyse data from a policy perspective;
- Limited access to existing agriculture data for users;
- Lack of metadata on agricultural statistics or quality indications

References

1. http://gsars.org/en/category/publications/handbooks-guidelines/
2. http://gsars.org/en/category/trainingmaterial/
3. http://gsars.org/en/category/regional/
4. http://gsars.org/en/e-bulletin-archive/
5. http://gsars.org/en/events/

COMUNIKOS (Communicating Uncertainty in Key Official Statistics): Application to measuring Well-being at local level for informed public policy

Tiziana Laureti[1], Dario Buono[2], Florabela Carausu[3]
[1]University of Tuscia, Department of Economics, Engineering, Society andBusiness Organization
[2]Eurostat-Methodology for Statistics
[3]Gopa Luxemburg

Abstract

The OECD initiative, to measure well-being, compiles internationally comparable indicators on welfare for the first time according to the recommendations made by the Stiglitz-Sen-Fitoussi Commission in 2009. 11 specific aspects of well-being are evaluated such as housing, income, employment, community, education, the environment, civic commitment, health, life satisfaction, safety, work balance - personal and two transversals aspects: sustainability and inequalities. The objective is to develop new measures to assess well-being and therefore it is recognized that the well-being of the population is not limited to the Gross Domestic Product (GDP) or economic indicators. Within the framework of this activity, the OECD launched, in 2014, the Regional Wellbeing Measurement project, which consists of a set of indicators of well-being and an analysis thereof, which allows su a better understanding of trends and of the welfare drivers in the 362 regions of the OECD member countries. From this initiative, the National Institute of Statistics and Geography of Mexico (INEGI) has developed a website (Site of Indicators of Wellbeing by Federal Entity) where there are twelve topics that cover material housing conditions and quality of life to measure the welfare of the 31 Mexican states and Mexico City. In this presentation, the experience of Mexico will be developed in the integration and application of the OECD welfare indicators as well as the main results found at the subnational level. By way of conclusion, we will also see the main challenges that the country has to reduce the disparities observed, this proposal of the OECD has the advantage of demonstrate how these indicators can be included in the design, and implementation of policies at all levels of government to improve the life of the people. In late 2018 Eurostat launched the COMUNIKOS project (COMmunicating UNcertainty in Key Official Statistics) which aims to conduct an in-depth methodological and empirical evaluation of a number of alternative approaches to measuring and communicating uncertainty as well to propose new methods and metrics in the official statistic setting. Although several statistical organisations have started to invest in identifying ways to

measure and communicate data uncertainty, this is only being done randomly. This may lead to a misinterpretation of the statistical results by the users.
In this paper we present an application of uncertainty in measuring economic well-being at local level, a crucial aspect that impacts both the design and evaluation of policies. We focus on the issue of measuring the accuracy of at risk of poverty rate (AROP) which is defined as the proportion of people with an equivalised total net income below 60% of the national median income and estimated from the European Union Statistics on Income and Living Conditions (EU-SILC) surveys. After reviewing the literature devoted to the computation of standard errors of traditional measures of monetary poverty and to the assessment of non-sampling errors, we estimate AROP standard errors for Italy taking into account the complex survey design and using linearization methods. Moreover, we provide estimates of standard error of AROP at regional level (NUTS-2) by also using equivalised income adjusted for price differences in order to better inform policy making. We use the 2017 wave of EU-SILC in which detailed information on the sample design are provided.

Keywords
Data uncertainty; well-being; poverty measures

1. Introduction

Despite the seminal studies of Kuznets (1948) and Morgenstern (1950) addressing the issue of uncertainties in economic statistics, the communication of data uncertainty is still a widely neglected problem. National Statistical Offices (NSOs) and other statistical agencies often publish official economic statistics as point estimates making little mention of uncertainty in the reported estimates. However, pretending an accuracy that is unrealistic may lead us to a misinterpretation of the statistical results and erroneous conclusions by the users. As Manski (2019) underlined, point estimates may be viewed as true but are not necessarily true. Thus, in the absence of agency guidance, some users of official statistics may simply assume that errors are small and inconsequential. Contrastingly, users of official statistics who understand that statistics are subject to error may conjecture the error magnitudes, thus misinterpreting the information that the statistics provide.

Therefore, in late 2018 Eurostat launched the COMUNIKOS project (COMmunicating UNcertainty in Key Official Statistics) which aims to conduct an in-depth methodological and empirical evaluation of a number of alternative approaches to measuring and communicating uncertainty as well to propose new methods and metrics to measure and communicate uncertainties for official statistics.

Measuring uncertainty is a complex and challenging task, which can involve the use of sophisticated statistical and econometric techniques and subjective judgement to quantify the data uncertainties.

Various frameworks have been proposed for defying and classifying uncertainty measures and their sources for economic statistics (Verma et al, 2010; Manski, 2015). Commonly, statisticians categorise uncertainties to reflect non-sampling and sampling errors. Sampling error is the most commonly reported measure of statistical uncertainty since this source can be quantitatively estimated for many sample surveys. Non-sampling errors apply to administrative records and surveys, including censuses, whereas sampling errors apply only to sample surveys. In principle, therefore, the total uncertainty associated with statistical output comprises both sampling error and non-sampling error (Mazzi et al, 2019). COMUNIKOS is designed to evaluate alternative ways of measuring and communicating data uncertainty specifically in contexts relevant for official economic statistics. Nevertheless, COMUNIKOS could be extended to the measurement of well-being which involves the use of economic and social statistics which are often provided by the statistical agencies without information on uncertainties.

In the framework of the OECD's Better Life Initiative, the economic well-being (material living condition) is identified as one of the pillars for understanding and measuring people's well-being following a multi-dimensional approach (OECD, 2013). Income and wealth are essential components of individual well-being since they allow people to satisfy their needs, to enhance individuals' freedom to choose the lives that they want to live, and to improve other important dimensions of well-being, such as life expectancy and educational attainments. The importance of economic well-being to overall well-being has been recognised by all the statistical agencies when producing framework for the measurement of individual well-being.

The growing attention to the measurement of well-being has led to progressively include well-being indicators in the policy agenda in various countries, including Italy where the Ministry of Treasury has started using well-being indicators for the evaluation of fiscal policies.

The micro approach for the measurement and analysis of economic resources available to the population refers to the study of poverty and its effect on different socio-economic groups within society. In the context of the Europe 2020 Strategy, poverty indicators play an essential role in informing and supporting responsible evidence-based policies towards the Union's key goals which relate to inclusive and sustainable growth and reduction of poverty, inequalities and social exclusion.

In 2010, the European Commission identified as one of the five headline targets of the Europe 2020 Strategy a 20-million decrease in the number of persons in or at risk of poverty and social exclusion (AROPE) by 2020. It is, therefore, essential that the indicators used for measuring poverty are of the necessary high quality, in particular in terms of their accuracy, timeliness, usability and their reliability. In this context, there is also an increasing demand for estimates of poverty and social policy indicators at regional or sub-national level (NUTS 2, and exceptionally NUTS 1) to be used for benchmarking and assessing the efficiency of regional policies (Piacentini, 2014).

Therefore, in this paper we present an empirical evaluation of uncertainty in measuring well-being at local level, a crucial aspect that impacts both the design and evaluation of policies. We focus on the accuracy of at risk of poverty rate (AROP) which is estimated from the European Union Statistics on Income and Living Conditions (EU-SILC) survey.

After reviewing the literature devoted to the computation of standard errors of traditional measures of monetary poverty and to the assessment of non-sampling errors, we estimate AROP standard errors for Italy taking into account the complex survey design and using various approaches, such as linearization and re-sampling methods. Moreover, we provide estimates of standard error of AROP at regional level (NUTS-2) in order to better inform policy making. We use the 2017 wave of EU-SILC for Italy in which detailed information on the sample design are provided by the Italian National Statistical Office (ISTAT) thanks to a research project carried out within the Dagum Inter-Universities Research Centre on Advanced Statistics for the Equitable and Sustainable Development.

2. Methodology

2.1 Measuring uncertainty for economic well-being

Over the last decades, there has been increased interest in comparative analysis of poverty and social exclusion in the EU both at national and at regional level. In 2004 Eurostat launched the EU Statistics on Income and Living Conditions in order to create a European standardized database to generate comparative measures of poverty and social exclusion among the Member States. In the framework of Europe 2020 strategy of the European Commission, estimates of poverty and social exclusion at sub-national regional level have become of central interest since local differences in poverty are essential for policy decisions and monitoring (Betti et al, 2012; Piacentini, 2014).

At the same time, Eurostat and a number of stakeholders are reflecting on ways to further improve the quality of statistics on income and living conditions based on EU-SILC survey, especially regarding their accuracy (Atkinson, 2017).

In this context, the issue of measuring the uncertainty of the poverty indicators that have been officially adopted by Eurostat has become paramount[1]. Indeed, it is crucial to take a measure of uncertainty into account when monitoring poverty in order to avoid misleading regional analyses and the consequent policy implications and outcomes.

Even if the importance of providing measures of uncertainty of poverty indicators has been widely recognized, NSOs are reluctant to communicate uncertainty measures of reported estimates in news releases while technical publications acknowledge that these statistics are subject to error by providing information on the methods used for estimation.

As EU-SILC is a sample-based survey, both sampling and non-sampling errors can seriously affect the accuracy of all estimates derived from this survey (Verma et al, 2010). However, the computation of standard errors for estimates of poverty indicators based on EU-SILC is a challenging task due to the complex sample designs employed by several EU countries which involves stratification, geographical clustering, unequal probabilities of selection for the sample units and post-survey weighting adjustments (re-weighting for unit nonresponse and calibration to external data sources for non-response adjustment of the initial design weights). Additional, EU-SILC has an important panel component with a 4-year rotational panel design in the majority of countries. Moreover, full documentation of the sample design and accurate sample design variables in the EU-SILC dataset are usually lacking, thus hampered research studies focused on exploring sampling error measures.

Several methods for estimating the variance of the poverty indicators have been discussed in the literature (Berger et at, 2017) These methods that can be classified into two approaches: ‘direct’ methods, which rely on analytic variance formulas through linearization (Alper and Berger, 2015), and ‘resampling’ methods, such as Jackknife repeated replication or boostrap, which consist of re-sampling a high number of ‘replications’ from the original sample in order to empirically derive a sampling distributions (Davidson and Flachaire, 2007; Verma et al, 2010). Contrastingly, standard error and confidence intervals of regional poverty indicator has been studied only in a limited number of papers (Verma et al 2017) although they are needed for regional (subnational) estimates for assessing and monitoring regional policies. In this context, the main difficulty arises from the smallness of regional samples in national surveys. Various methods could be used to overcome this issue as illustrated by Verma et al 2017. However, it is worth noting that whichever approach is used, in order to obtain reliable standard error estimates it is essential that the sample

[1] By referring on the content of intermediate and final EU-SILC Quality reports, Commission Regulation (EC) No 28/2004 of 5 January 2004 requires that countries should provide estimates of standard error along with the EU-SILC main target indicators.

design, the weighting procedure, the imputation schemes and the nonlinear form of survey estimators should be reflected in the calculation of standard errors and confidence intervals (Goedemé, 2013).

a. *Estimating standard errors: An Empirical analysis for Italian regions*

In Italy, EU-SILC is conducted by ISTAT to produce estimates of the Italian population living conditions at national and also at regional level (NUTS-2). In the design of the EU-SILC survey, regions are planned domains for which estimates are published but without confidence intervals (ISTAT, 2018). The regional samples are based on a stratified two-stage sample design: in each province, municipalities are the Primary Sampling Units (PSUs), while households are the Secondary Sampling Units (SSUs). The PSUs are stratified according to administrative regions and population size while SSUs are selected by using systematic sampling. We calculate confidence intervals for the AROP for Italian regions in 2017 (NUTS-2) using data from IT-SILC provided by ISTAT for carrying out research projects in collaboration with the Dagum /Tuscan Universities Research Centre on Advanced Statistics for the Equitable and Sustainable Development.

AROP is a complex statistic since it is based on a poverty threshold computed from the median of the income distribution, that is: $AROP = P(X < 0.6 \cdot q_{0.5}) \cdot 100$. Therefore, the at-risk-of-poverty threshold (ARPT) needs to be estimated first, which is set at 60% of the national median equivalized disposable income, $\widehat{ARPT} = 0.6 \cdot q_{0.5}$. Then the AROP rate is defined as the proportion of persons with an equivalized disposable income below the ARPT: $\widehat{AROP} = \frac{\sum_{i \in I < \widehat{ARPT}} w_i}{\sum_i w_i} \cdot 100$

Consequently, there exist two main sources of variability: one is due to the estimated threshold and the other one comes from the estimated proportion given the estimated threshold (e.g., Berger and Skinner 2003; Verma et al 2012). We used a generalized linearization method based on the concept of influence Function (Deville, 1999) which allows us to deal with nonlinear statistics for which the Taylor method cannot be used. This approach does not involve more calculations.

We linked household and personal data (H, R, P D files). Household income is equivalised using the modified OECD equivalence scale by assuming that the living standard of all household members is the same. Confidence intervals are estimated using linearization on the basis of complete sample design information regarding “Primary strata”, “PSUs”, “SSU”. The estimates are obtained using R package ***laeken*** and the DASP module. Since a region coincides with a sample “design domain” and the regional AROP depends only on units within region, variance computations can be performed for each region in the same way as at the national level.

Since the poverty threshold chosen can influence the resulting poverty rates when there is a high degree of disparity between the units of analysis, as in the case of Italy, we estimated 95% confidence intervals for AROP using two different procedures based on a national poverty line estimated from the sample by means of: a) equivalised household income and b) equivalised household income adjusted for price differentials among regions using as proxy of the overall sub-national Spatial Price Indexes (SPIs) the “Food products” SPIs calculated using scanner data from modern retail chains. These SPIs are obtained within an ISTAT research project for computing sub-national SPIs based on scanner data and CPI data (Laureti and Polidoro, 2018). Even if SPIs for food consumption aggregate represents only a part of the total household consumption expenditure it may be interesting to analyse what happens when price dimension is included into AROP standard error measurement. Indeed, in this case an additional source of uncertainty is introduced. Further research will be devoted to this issue within the COMUNIKOS project.

Another critical aspect in measuring uncertainties in poverty indicators is how to communicate them in a “comprehensive” way, in terms of capturing fully the uncertainties, but also in a “understandable” way so that different users and readers of these data correctly infer and interpret the uncertainties communicated to them. Increasing attention has been paid to this aspect in literature (Spiegelhalter et al., 2011; van der Bles et al., 2019). COMUNIKOS aims also to carry out a detailed investigation of the pros and cons of communicating uncertainties to users of official statistics by considering appropriate tools for measuring and disseminating data uncertainties.

3. Results

With the aim of providing an idea of the role visualization when communicating uncertainty in poverty indicators, we explore various ways for displaying uncertainty using bar charts (van der Laan, 2015). Figure 1a depicts the point estimates for AROP without information on standard errors, as it is usual practice, while Figure 1b shows only 95% confidence intervals for Italian regions. Figure 2a and Figure 2b show AROP point estimates together with 95% confidence intervals for Italian regions. In these figures bar chart with error bars are compared with bar chart with cross bar. While the first chart puts a visual focus on the point estimate the cross-bar chart puts more emphasis on the uncertainty measure. Crossbar chart seems to allow a clearer interpretation of confidence intervals by putting users' attention on both lower and upper limits.

Figure 1 AROP estimates without uncertainty measure and AROP with 95% confidence interval for Italian regions (using national poverty line)

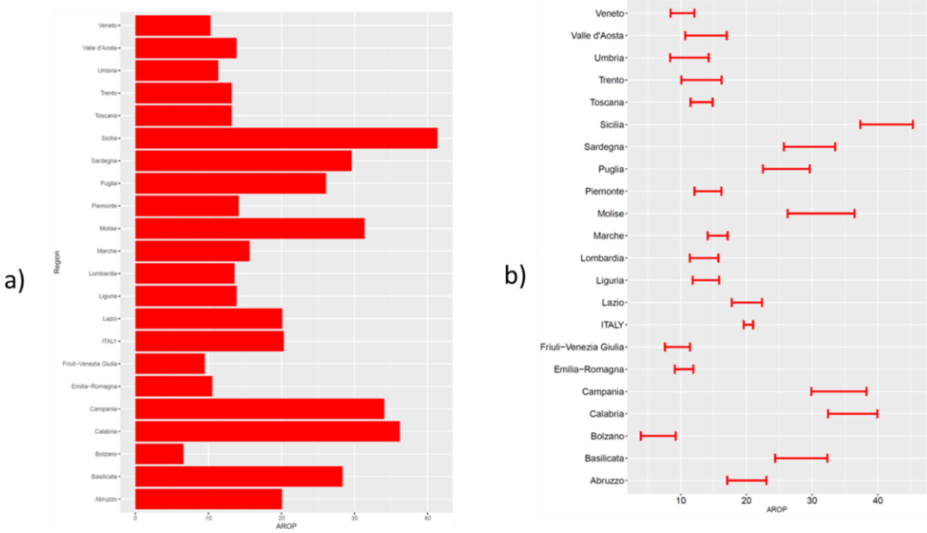


Figure 2 AROP with 95% confidence interval for Italian regions (using national poverty line)

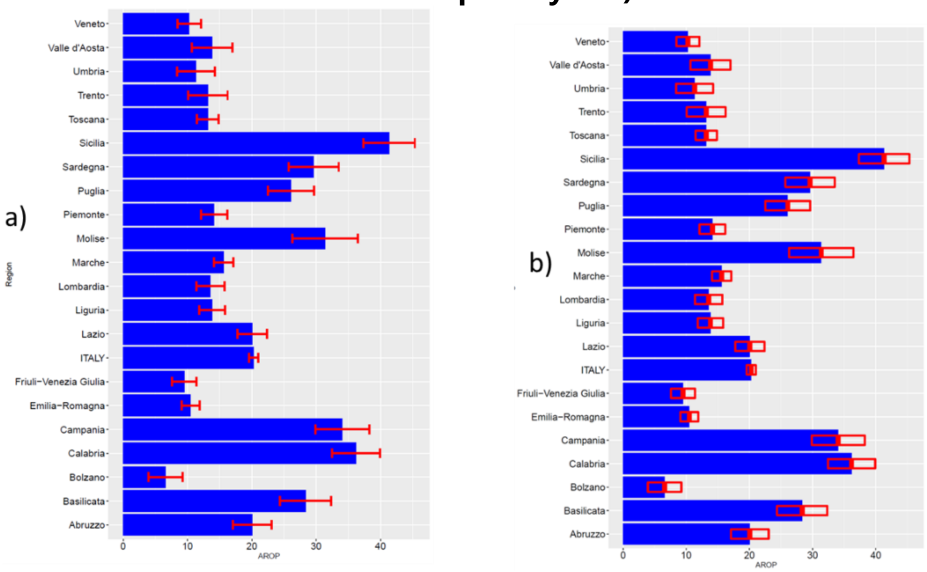
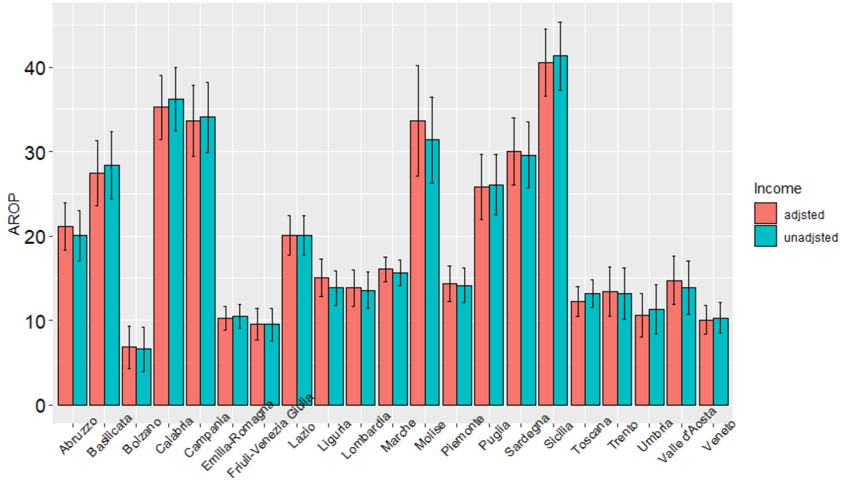


Figure 3 shows the results obtained when equivalised household income is adjusted using "Food products" SPIs as proxy of price differentials among regions.

Figure 3. AROP with 95% confidence interval for Italian regions (using adjusted and unadjusted equivalised income)



When adjusted income is used, point estimates of AROP assume higher values for some Italian regions, such as Abruzzo, Molise and Liguria and lower values for other regions, including Basilicata, Toscana, Calabria and Sicilia. Uncertainties of AROP increase for some regions (Molise, Puglia, Liguria) while decrease for Umbria, Valle d'Aosta, Abruzzo. Caution is required when interpreting these results since they may be influenced by the characteristics of the modern retail trade which is not uniformly distributed across Italian territory in terms of types of retail chains and market share.

4. Discussion and Conclusion

Using data from 2017 EU-SILC with detailed information on sampling design and variables, we estimated sampling errors for AROP for Italian regions using linearization method and taking into account price differentials as measured by a Food products SPI constructed using scanner data. The results seem to suggest that data uncertainties present in AROP need to be provided by NSOs for informed public policy. The measure of uncertainty is influenced by the introduction of price statistics which also decrease heterogeneity across Italian regions. These results suggest interesting lines for future research on the measurement of uncertainty in economic well-being and in price statistics.

References

1. Alper, M. O., & Berger, Y. G. (2015). Variance estimation of change in poverty rates: an application to the Turkish EU-SILC survey. Journal of Official Statistics, 31(2), 155-175.
2. Atkinson, A. B., Guio, A. C., & Marlier, E. (Eds.). (2017). Monitoring social inclusion in Europe. Luxembourg: Publications Office of the European Union.
3. Berger, Y., Osier, G. and Goedemé, T. (2017), Standard error estimation and related sampling issues, in A. B. Atkinson, A.-C. Guio and E. Marlier (eds.), Monitoring social inclusion in Europe, Luxembourg: Publications Office of the European Union, pp. 465-478
4. Davidson, R., & Flachaire, E. (2007). Asymptotic and bootstrap inference for inequality and poverty measures. Journal of Econometrics, 141(1), 141-166.
5. Deville, J. C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. Survey Methodology, 25, 193-203.
6. Goedemé, T. (2013). How much confidence can we have in EU-SILC? Complex sample designs and the standard error of the Europe 2020 poverty indicators. Social Indicators Research, 110(1), 89-110.
7. Laureti T. and Polidoro, F. (2018) Big data and spatial comparisons of consumer prices, Proceeding of 49th Scientific Meeting of the Italian Statistical Society, Palermo, Book of short Papers SIS 2018, Pearson, ISBN-9788891910233
8. Manski, C. F. (2015). Communicating uncertainty in official economic statistics: An appraisal fifty years after Morgenstern. Journal of Economic Literature, 53(3), 631-53.
9. Mazzi GL, Michell J. and Carausu F. (2019), Data uncertainties: their sources and consequences, Deliverable task 1, COMUNIKOS.
10. OECD (2013), "Economic well-being", in OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth, OECD Publishing, Paris
11. Piacentini, M.,(2014). Measuring income inequality and poverty at the regional level in OECD countries. In: Proceedings of the OECD Statistics Working Papers, 2014/03.OECD Publishing
12. Spiegelhalter, D., Pearson, M., and Short, I. (2011). Visualizing uncertainty about the future. Science (New York, N.Y.), 333(6048):1393–400.
13. van der Bles, A. M., van der Linden, S., Freeman, A., Mitchell, J., Galvao, A. B., Zaval, L. & Spiegelhalter, D. (2019), Communicating uncertainty about facts, numbers and science, Royal Society Open Source 6, 1-42.

14. van der Laan, D., de Jonge, E., & Solcer, J. (2015). Effect of Displaying Uncertainty in Line and Bar Charts-Presentation and Interpretation. In IVAPP (pp. 225-232).
15. Verma, V., Betti, G. & Gagliardi, F. (2010), An assessment of survey errors in EU-SILC,Eurostat: Methodologies and working papers.
16. Verma V., Lemmi, A. Betti, G. Gagliardi, F. Piacentini M. (2017), How precise are poverty measures estimated at the regional level? Regional Science and Urban Economics 66,175–184

The OECD approach: The case of Mexico

Oscar Gasca Brito
Instituto Nacional de Estadística y Geografía, Aguascalientes, México

Abstract

The OECD initiative, to measure well-being, compiles internationally comparable indicators on welfare for the first time according to the recommendations made by the Stiglitz-Sen-Fitoussi Commission in 2009. 11 specific aspects of well-being are evaluated such as housing, income, employment, community, education, the environment, civic commitment, health, life satisfaction, safety, work balance - personal and two transversals aspects: sustainability and inequalities. The objective is to develop new measures to assess well-being and therefore it is recognized that the well-being of the population is not limited to the Gross Domestic Product (GDP) or economic indicators. Within the framework of this activity, the OECD launched, in 2014, the Regional Wellbeing Measurement project, which consists of a set of indicators of well-being and an analysis thereof, which allows su a better understanding of trends and of the welfare drivers in the 362 regions of the OECD member countries. From this initiative, the National Institute of Statistics and Geography of Mexico (INEGI) has developed a website (Site of Indicators of Wellbeing by Federal Entity) where there are twelve topics that cover material housing conditions and quality of life to measure the welfare of the 31 Mexican states and Mexico City. In this presentation, the experience of Mexico will be developed in the integration and application of the OECD welfare indicators as well as the main results found at the subnational level. By way of conclusion, we will also see the main challenges that the country has to reduce the disparities observed, this proposal of the OECD has the advantage of demonstrate how these indicators can be included in the design, and implementation of policies at all levels of government to improve the life of the people.

Keywords

Wellbeing; territorial indicator; Mexico; regions

1. Introduction

Can GDP growth measure the well-being of a population? Are economic indicators alone enough to analyze the progress of a nation or is it just one face of the coin? In a global context in crisis, in 2008, the President of France instructed JP Fitoussi, S.Stiglitz and A. Sen to establish a commission on the measurement of economic development and social progress whose main objective was to determine the limits of GDP as an indicator economic and social progress.

The conclusions were overwhelming: welfare is multidimensional, the economic indicators used, alone, are insufficient to explain population social progress and economic growth in an increasingly changing technological context. The commission determined that in order to delimit the notion of welfare, in principle, at least, the following dimensions should be apprehended simultaneously[1]:

- Material living conditions (income, consumption and wealth)
- Health
- Education
- Personal activities and within them work activities
- Participation in political life and governance
- Social ties and relationships
- The environment (the present and future state)
- Insecurity, both economic and physical insecurity

The following recommendation is released[2]: "the quality of life depends on the objective conditions in which people find themselves and their dynamic capacities. It would be convenient to improve the statistical measures of health, education, personal activities and environmental conditions. In addition, a particular effort should be given to the conception and application of solid and reliable tools for measuring social relations, participation in political life and insecurity, a set of elements that can be shown to be a good predictor the satisfaction that people get from their life”

This assessment was reaffirmed at the UN, at the World Summit Meeting on Sustainable Development, where it was recognized the urgent need to complement the measurement of gross domestic product with a variety of indicators that allow integrate both the objective and subjective dimension of economic and social development.

The territorial indicators of well-being developed in Mexico are part of this complex context, in which international and national community was looking for indicators with the level of disaggregation needed to measure both objective and subjective well-being in a complicated country with so many disparities like is Mexico.

[1] Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2009). Report by the commission on the measurement of economic performance and social progress. Paris. http://www.stiglitzsen-fitoussi.fr/en/index.htm

[2] Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2009). Report by the commission on the measurement of economic performance and social progress. Paris. http://www.stiglitzsen-fitoussi.fr/en/index.htm

2. Methodology

In this international context, the Organization for Economic Cooperation and Development (OECD), of which Mexico is a member, echoed that proposal and in 2011 launched the initiative for a Better Life, which is composed of two major elements:

- How is life?: a report that for the first time, brought together eleven specific aspects of well-being such as housing, income, employment, community, education, the environment, civic engagement, health, satisfaction with the life, safety, balance between life and work, as well as two transversal dimensions: sustainability and inequalities for all OECD countries.
- The Better Life Index: an online tool that proposes to the citizens of the different member countries to join the debate on social progress and to qualify each one of the eleven themes and thus create their own perspective to measure and compare welfare between countries. Each of the 11 topics is based on one or more available indicators and within each theme; the indicators are calculated with similar importance level. The indicators were chosen based on various statistical criteria such as relevance (apparently depth political relevance), quality of the information (prediction validity, coverage, punctuality, comparability between countries, etc.) as well as consultations with member countries of the OECD. The following indicators composed each themes:

Table 1: OECD themes and indicators

Themes	Indicators
Housing	Housing expenditure Dwellings with basic facilities Rooms per person
Income	Household financial wealth Household net adjusted disposable income
Jobs	Job security Personal earnings Long term unemployment rate Employment rate
Community	Quality of support network
Education	Years in education Student skills Educational attainment
Environment	Water quality Air pollution

Themes	Indicators
Civic engagement	Voter turnout Stakeholder engagement to developing regulations
Health	Self-reported health Life expectancy
Life satisfaction	Life satisfaction
Safety	Homicide rate Feeling safe walking alone at night
Work-life Balance	Time devoted to leisure and persona care

How you can see the Better life Index brings together many indicators, expressed in different units (dollars, years, percentage, etc.) so its standardization is necessary to facilitate comparison between countries. This process is performed according to a standard formula that converts the original values of the indicators into numbers that vary in a range of 0 (for the worst possible result) and 1 (for the best possible result).

Based on these indicators, the user can generate a “flower”, assigning the importance it considers fair to each theme where each “flower” represents a country, each petal one of the 11 themes and the length of the petal represents the country's rating on the subject, while the width represents the importance that the user assigned.

Parallel to international work to approach a measurement, both objective and subjective, of welfare, in May 2011, takes place in Mexico City, the Latin American Conference for the measurement of welfare and the promotion of the progress of societies whose objectives were:

- Deepen the debate on how to measure social progress and well-being in Latin America;
- Improve the relevance of current measurements and analysis to address fundamental policy issues related to progress and social welfare, and;
- Achieve concrete results, establish frameworks and open new routes for future work[3]. The main conclusions of that Conference were:
- The measurement of welfare goes beyond money and GDP, and requires the consideration of objective and subjective dimensions. The measurement of well-being must focus on individuals and households.
- In Latin America, the important dimensions of well-being include health, education, working conditions, housing, economic situation, interpersonal relations, availability of free time, access to social protection, effective citizenship, law enforcement and ethnic and gender equality.

[3]INEGI (2011), Conclusiones de la Conferencia Latinoamericana para la Medición del Bienestar y la promoción del progreso de las sociedades. http://mfps.inegi.org.mx/Default.aspx

At that time, the Instituto Nacional de Estadística y Geografía (INEGI) joins the OECD proposal and undertakes the design of a tool on its website that would allow viewing the welfare indicators defined within the OECD and adapted to both the availability and information needs of the country. The ultimate goal of this collaboration is to have a significant set of welfare indicators for citizens as well as those responsible for public policies to use to design public policies and improve their results. These indicators will allow evaluating the living conditions in the Mexican states in the different dimensions of well-being, covering both aspects of material conditions as well as the quality of life.

In order to fulfill this mission, a series of meetings with the OECD and the signing of an agreement between both institutions are undertaken whose purpose is to build a system of territorial indicators of welfare and thus have an input of public policies at the local level, and give bases to the governments to undertake public policies that allow to improve the life of people in a tangible way. In this time, Mexican states have demanded to INEGI a system of indicators to be comparable, transparent, updated with their metadata, in a timely manner that allows them to compare and to contrast successful public policies in well-being. INEGI took the decision to implement this initiative in Mexico and suggested the governor of the state of Morelos to conduct a pilot test, together with United States, Italy, Denmark, UK, and the Netherlands. The result of this test was very successful, and Morelos continued working with the OECD until 2018, at the end of the government period. After this, INEGI decided to bring all the Mexican states, not only with a document that let us know the situation of each of the states and compare them but through a web site that allows these indicators to be permanently updated, with time series, with different way of viewing these indicators with their metadata. In such a way, that they can be consulted at any time and not lose its validity. Something very important to say in this project is that it was not just a work of the OECD and INEGI each of the Mexican states were participating from the beginning. So that, it is not a project with a focus on a product, but on users, because they participated in the definition of indicators that could be more useful for all the Mexican states.

For the case of Mexico, the work group defined 12 themes with their respective indicators grouped
as follows:

Table 2: INEGI themes and indicators

Themes	Indicators
Accessibility to services	Health serviles access Dwellings with internet Dwellings with basic facilities
Community (social relationships)	Quality of support network
Education	Levels of education School dropout Average years of schooling
Life-Work Balance	Time devoted to leisure Employees working very long hours (over 48 hours)
Income	Gini Index of disposable income per capita household Equivalent disposable income of households Poverty population (%) Extreme poverty population (%)
Environment	Air pollution Waste disposal
Civic engagement	Civic and political participation Voter turnout Confidence in the application of the law Perception of absence of corruption in the system
Health	Life expectancy Self-reported health Obesity rate Maternal mortality rate Infant mortality rate Reason for maternal mortality (deaths per 100 thousand livebirths)
Life satisfaction	Life satisfaction
Safety	Homicide rate Confidence in the police Perception of insecurity Crime incidence rate
Jobs	Occupancy critical conditions rate Labor informality rate Unemployment rate Economic participation rate
Housing	Rooms per person Percentage of housing with roofs of precarious

In Mexico, the level of well-being of people can be very different depending on the place considered. The disparities among the Mexican states in matters of health, safety, education and civic participation are among the five largest of the OECD countries. As in other OECD regions, several Mexican states have launched initiatives to improve the welfare of their citizens, developing metrics to monitor the progress of society and the results of its policy. A crucial asset to help make these initiatives effective is to develop a solid system of welfare indicators at the state level, which makes it possible to compare living conditions between states and internationally and monitor how this condition evolves.

At the regional level, in August 2013, the state of Morelos, in collaboration with the OECD and INEGI, undertook the project "Use of well-being indicators for the design of public policies: State of Morelos, Mexico", which represents a case of pilot study, applicable to other Mexican states, and which provides a welfare analysis at the federative level based on an international framework, as well as tools that allow knowing how to use welfare measures to improve the results of public policies. This collaboration consisted, within the framework of the initiative "What is life like in your region?", In identifying the appropriate dimensions and welfare measures for the entity, through a process of consultation with relevant local actors, and the identification of the information available for the development of welfare indicators, as well as information needs.

3. Results

As a result of this collaboration, the OECD delivered the study to the state of Morelos: use of welfare indicators for the design of public policies: state of Morelos, Mexico, which provides a diagnosis of the welfare state, as well as a battery of 19 indicators classified in 10 dimensions, according to the methodology implemented by the OECD for the study as well as a series of recommendations for the entity. Consequently, the state of Morelos was the state that accompanied the INEGI in the definition of the battery of indicators based on the experience acquired with the OECD.

In Morelos, the State Development Plan (PED) 2013-2018 was developped through a Citizen Consultation Process and had the main objective of building a society of rights to improve the welfare of citizens. In order to follow up on the objectives of the PED, 72 indicators were included in the five strategic axes of the Plan:

- Safety and Justice,
- Social cohesion and citizenship,
- Competitiveness and Innovation,
- Sustainability of the environment,
- Transparency and democratic participation

The objectives were: to promote local discussion on the most relevant welfare dimensions at the regional level and in the daily life of its inhabitants, as well as to identify the welfare metrics appropriate for the local context, which will contribute to improving the design, monitoring and the evaluation of public policies, as well as the rendering of accounts. In order to meet these objectives and undertake a public policy based on evidence, a series of meetings were organized with state agencies (education, environment, social development, entrepreneurs, security and justice, most important municipalities of the state), the Institute of Development and Municipal Strengthening of the State of Morelos (IDEFOMM), the INEGI and the OECD. In this consultation process, the data available at the state level was identified for the development of welfare indicators, as well as information needs.

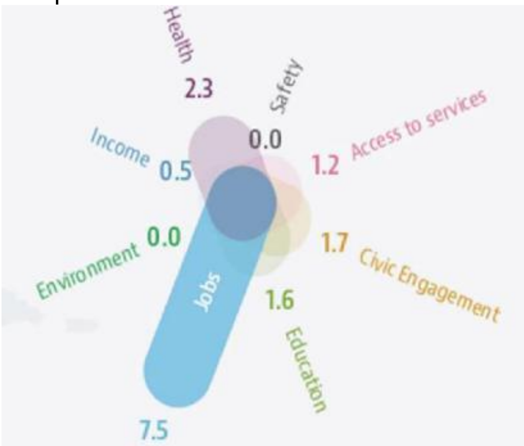
Following the methodology of the OECD, eight themes and the following indicators were identified:

Table 3: State of Morelos themes and indicators

Themes	Indicators
Education	Dropout rate in secondary Mathematics score
Jobs	Female participation rate Employment rate
Income	Available income per household Income gap between low income and average
Safety	Victimization rate Perception of security
Health	Life expectancy Overweight and obesity in the population of 12 to 19 years.
Environment	Air quality % adequate waste disposal
Civic engagement	Transparency Index
Accessibility to services	People with access to healthcare over the total population. Proportion of the population without access to public transport.

For Morelos, the following graph was generated, which allowed the local authorities to focus on the issues where there was more delay in the population, as well as comparisons within Mexico and around the regions of the OECD member countries.

Graph 1: Better life “flower” for Morelos



The results showed that the state of Morelos is among the top ten in the dimensions of employment, health and civic participation, and above the national average in education. It also allowed finding the most important challenges for state public policy: to improve the level of income and environmental quality. The state focused its welfare agenda on the following policies:

- Income: Reduction of income inequalities; attraction of foreign direct investment (FDI); social inclusion and cohesion and poverty reduction.
- Employment: reduction of labor informality to avoid a negative impact on access and the quality of jobs.
- Housing and access to services: control the urban growth, diversify housing options, improve urban mobility and increase the efficiency of land use. In addition, improve accessibility to certain services such as internet, public transport, schools and health centers, as well as, housing connections to basic services (water, sewerage, waste discharge).
- Health: focus on the results of health indicators, such as the incidence of diabetes or obesity, in order to better control health conditions.
- Education: Improvements in the coverage and quality of education. Increase the coverage of high school and college, and monitor the skills and achievements of students.
- Environment: Improve and control air quality. Fight against pollution and water scarcity, and focus on the proper disposal of solid waste.
- Security: measurement of objective and subjective security to improve security personnel and the level of confidence of citizens in the authorities and the police.

4. Discussion and conclusions

Well-being must be measured where it matters. The characteristics of a region or city can strongly affect the well-being of the people who live there. Improving people's lives requires that their region be a better place and the availability of solid welfare measures is crucial for such improvement. In this sense, indicators at the national level on material conditions and quality of life can be misleading and hide important disparities between regions and cities within the same country[4]. In addition, many of the policies that most directly affect people's lives are local or regional and, therefore, require indicators at this scale.

However, we still have a way to go and in an international context where we can access more and more information more quickly, we have several challenges such as statistical institutes that resolve:

- On the one hand, how to make the SDGs, OECD indicators and national initiatives compatible? Users can be saturated with indicators and information as well as the level of disaggregation of each one. Among all the initiatives, there may be more than 80 indicators without counting territorial disaggregation, by sex, by age, etc. It is imperative for international as well as national organizations to design a simple and user-friendly indicator system for the population as a whole, regardless of the level of disaggregation required.
- On the other hand, we have to keep focus on users, this factor is very important to better understand their information needs or if additional indicators are required. The timely updating of indicators, since the timed release of the data is the success factor of the information generated. Having internationally comparable indicators, in order to compare the different public policies among OECD countries and their regions in terms of Well-being.
- Finally, to have more information at urban and rural level, given the importance of having the decision-making increasingly accurate and closer to the people. Think globally; act locally, but for this you need to have that information. For this, it is necessary to exploit other sources of information such as administrative records, big data, etc. to be able to measure well-being on a smaller scale in order to help local government with their policies.

[4] OECD (2014) Guidelines on Measuring Subjective Well-being http://www.oecd-ilibrary.org/economics/oecd-guidelines-on-measuring-subjective-well-being 9789264191655-en

A discussion about the teaching of statistics at all levels in an Argentinian province

Terán, Teresita
Veterinary Sciences University – National University of Rosario. Rosario. Santa Fe Province. Argentina

Abstract
The dynamic nature of statistics requires a continuous investigation of contents and teaching methods at different educational levels. Technology, for its part, is currently influencing what and how to teach; producing a constant revision and analysis of contents, which are conditioned by the previous hypotheses that determine learning. What kind of teaching practices on probability and statistics should be implemented to promote statistical knowledge? how do we foster the critical statistical sense on our students' learning?, what contents on probability and statistics could promote intra-mathematical mainstreaming and interdisciplinarity fostering the work based on projects? Nowadays, in this ever-changing and unpredictable society, statistical education is taking place in a new social context. This one is influenced by a movement of teaching reform of mathematical sciences in general.

Keywords
Statistics, Teaching, Educational level

1. Introduction

As teachers we should reflect upon the best way to prepare young people and what the contents and subjects we should teach are, since the contents that we consider important today, most probably will not be so in the coming years. Let's remember the transitions from manual calculations to calculators and then to computers. Another question arises then: What are the competences students need for their future performance, not only for their daily life but also for their professional life? On this respect Moore claims: "The reason why there should be substantial changes on statistical teaching is because there is a strong synergy between the contents, pedagogy and technology. Statisticians who teach at introductory levels should familiarize themselves not only with investigations on teaching and learning, but also with changes on educational technology. The spirit of instruction upon contemporary statistics will have to be very different from the traditional one which made emphasis on explanations and on probability and inference."

Probability and statistics are two very important components in our culture and in many of our specific sciences. They should be part of the basic cultural knowledge of every citizen in our society. This is something on which every educational system seems to agree on. Holmes (1980) claims that statistics is a part of the desirable general education provided to future adult citizens, who need to acquire the ability to read and understand statistical charts and graphs that appear most of the times on media. In order to be able to move around successfully in today's world, connected by telecommunications and socially, economically and politically interdependent, it is necessary to interpret an extended range of information on the most varied topics.

Ottaviani (1998) considers that the study of statistics helps personal development, fostering critical reasoning, based on judgement of objective evidence; students should be able to use quantitative data to understand it and explain it in context. A successful scientific background requires instruction on probabilistic and statistical thinking.

These recommendations have motivated the idea of statistics to become part of the syllabus more and more. Statistical reasoning is a major component of essential learning, as suggested by Wild y Pfannkuch (1999) who make emphasis on the integration of statistics and context.

2. Development

The approach to the introduction of statistical instruction on syllabuses at different levels of education has already been stated; we will now turn the focus on Argentina, showing the remarkable changes that have been registered.

Until 1993, at initial and primary level of education no topic of statistics had been covered. At higher levels some topics on descriptive statistics were covered only as part of math syllabuses at 4th year of education. They were rarely taught by teachers since they had not had statistics on their teacher training courses. Only a few schools that were dependent on universities, imparted statistics as a curricular subject. On this respect, in 1993 when the Federal Law of Education Nº 24.195 was enacted, contents of statistics and probability started to be taken into consideration to be included as part of the Basic Common Contents for the different levels and cycles of the new academic structure.

Statistics started to be considered on the different levels contemplated by the Federal Law of Education (1993) as a segment of math syllabuses.

The new National Educational Law changed the structure of the National Educational System. It comprises 4 levels: initial, primary, secondary and higher education. Each jurisdiction embraced the general guidelines and made adjustments based on their regional characteristics. From here onwards, we will present and analyze the jurisdictional approach implemented in Santa Fe province.

Initial Education entitles a place of subjectivity and social responsibility, it implies a period of time and space that receives and embraces boys and girls from three to five years of age included, as a first encounter with the academic organization of the Educational System. As a constructor and transmitter of culture, Initial Education will allow children, through games, to explore and appropriate the world that opens up beyond their affective core and family learning; developing personality foundations in a period of conquests with such an intensity that it will never be the same. In such environments it is expected that children learn through a process of construction of meaning and sense of things, persons, nature, reality, life itself. Initial Education fosters the development of different languages of expression: expressive, physical, communicative, symbolic, playful, cognitive, ethic, affective, metaphoric, logic, imaginative, relational, with big respect to whoever the learner is. At this stage, we can observe a child who is powerful, energetic, and who has the ability of self-construction while they transform the world, with the strength of who grows and wishes to grow, with the wealth of curiosity and surprise that leads themr into asking the whys behind everything, at the peak of their senses, original, creative, with the desire to learn and live. A child who holds values, builds solidarity, and is available to the new and different.

It is here where statistics play a fundamental role. The big issue is that initial education teachers do not count with statistical knowledge to bring to their classrooms. Nowadays, they have started to deliver courses through the school teachers' union AMSAFE where they instruct teachers, and that knowledge is taken to classrooms through games. Children learn by playing. In this process, teamwork, communication and feedback improve learning, so much so as the formulation of questions where the why is highlighted.

Primary Education has the aim of providing an integral, basic and common instruction. This is compulsory and constitutes a pedagogic and organizational unit intended to develop of children from six years of age. Its purpose is to provide an integral, basic and common instruction and it aims to guarantee every child the access to a set of common knowledge that will allow them to participate fully in their family, school and social life. From six years of age boys and girls develop at school a wide range of mental and physical constructions: they invent, practice, feel, express with their bodies, with words, symbols, and art.

In this seven-year journey, letters and numbers, texts and reflexive situations, life in nature, history, listening to other people's opinions lay down the ways each boy and girl chooses for themselves.

Statistical contents are presented in primary level to be developed in an increasing way from first to seventh grade, but here the problem of not having enough teacher knowledge is presented, even though notions of statistics have already been included in their teacher training courses syllabuses.

Through a survey organized by the Ministry of Education it can be observed that statistical contents at this level are almost inexistent; it is claimed that since they are mostly at the end of the curricula, they are hardly ever covered. This is the second big failure observed.

Secondary Education is divided into two cycles: A Basic cycle, common to every orientation and an Oriented cycle, specific to different areas of knowledge, the social world and work. Secondary Education is compulsory and constitutes a pedagogic and organizational unit devoted to adolescents and young people who have completed the level of Primary Education. Secondary Education, including all its modalities and orientations, aims at enabling adolescents and young people to their full commitment as citizens, for their work and for their future studies. Young age entitles that adolescents face the challenge to build their own life projects. At this stage, young people search, rehearse, take and leave, try their limits, define their tastes and ideas. At Secondary Education, within the math syllabus, a set of knowledge on statistics and probability is imparted. Historically, mathematical thinking has been based on a deterministic idea which has excluded the intervention of those variables which gave place to unpredictable processes from the solutions that math provided, an idea that has been reinforced from school math. It is important to solve problems that allow the recognition and use of probability as a way to quantify uncertainty.

Resources used on media to describe information have a great mathematical foundation and citizens should be prepared to understand what they receive and make decisions based on that. In this respect, we can consider the most varied sections on newspapers where we can find expressions such as average value, tendencies, estimates, chance that belong to the jargon of Statistics. This situation makes the mathematical work in relation to the newspaper very important, because apart from the mathematical thinking, it comprises the ethics behind teaching students to discriminate the way we inform.

Likewise, it is important that students collect or find the information on databases. This will be the starting point for the formulation of questions or hypotheses about the problems to solve. It becomes necessary to organize the information into charts or graphs, by the ICT, and based on the analysis of the data gathered, to select the statistical measures that can be more representative.

Probability like statistics are concept integrators, which allow their treatment with contents from other axis, like sides of the same mathematical work, which include not only deterministic thinking but also random thinking. Higher Education comprises universities and university institutes, state-owned or private authorized, in accordance to the established denomination in Law Nº 24.521.

If we analyze statistics as a subject, we can see that it is present on the syllabuses of most non-statistical courses imparted at university, as a tool where its role is fundamental.

Then, it is necessary to pose the question if its presence in the respective syllabuses is adequate and if the proposed contents in each case match the requirements of professional performance.

In this respect, it is essential that when we determine the contents of each syllabus, these are related to the basic statistical knowledge that backs up the need to satisfy the requirements of other topics, many of them very important to the proposed course. We have to become conscious of the fact that reality is in constant change and that university has to move forward with these changes.

Because of this, we have to consider if the current contents in the statistics' syllabus of the different courses match the needs of the future professional. It is intended that the future graduate manages statistical language and knows which statistical methods can be applied.

In this respect, in the IV International Conference of Statistics Teaching (ICOTS-4) in Marrakech 1994, the Spanish speaking group on "The statistical syllabus for 2000" formed by 28 teachers representing four countries: Argentina, Colombia, Spain and Mexico gave some advice on the teaching of statistics in other courses. Among them:

- It is better to leave the statistical courses for the last semesters, when the students need that knowledge for their thesis.
- These courses should be articulated with courses such as methodology of investigation, for instance.
- They should not have too many topics; just a few are necessary and these should be interesting and illustrative, so that they could act as a motivation for future studies.
- Make use of everyday situations.
- Focus teaching on increasing awareness of the need to consult statistical professionals from the beginning of investigation, when the objectives are outlined and hypotheses are posed, and not only when the data has been collected.

Nowadays, this advice is not being taken into account in most universities of Santa Fe province, since statistics as a subject is imparted in the first or second year of the course and it is a four-months subject where even though descriptive and inferential statistics are included in the syllabus, only some topics are taught due to lack of time.

3. Conclusion

It is undoubtable that the twentieth century has been the century of Statistics, which has come to be considered as one of the fundamental methodological sciences and a base of the experimental scientific method. Teaching statistics, however, does not seem to follow the same path, even though Batanero highlights that in the last decade there has been a particular increase of the use of statistical ideas in different disciplines. This can be observed in scientific magazines and in the increasing implication of statisticians in the interdisciplinary work teams and in many cases misapplied. We think that this indicates the existence of an educational problematic that has its roots in the incorporation of statistics from school. This is not yet a fact. Even though statistics is included in Primary and Secondary Education syllabuses, teachers generally leave this topic for the end of the programme and many times it is overlooked. Students reach university without the basic knowledge and it is necessary to start the year repeating the contents of descriptive statistics and calculation of probabilities that should have been acquired at school.

University teachers, who should try to reach statistical inference – at least at the beginning- since this is the topic that will be most helpful for students, should speed up their explanations, suppress practical activities, and a big part of demonstrations or reasoning that could lead to the students´ better understanding of the methodology of statistics. The students cannot assimilate the content in such a limited period of time and only achieve learning content by heart which will be unable to apply later on in their professional life. All these nuisances are worsen by the big number of students in the classrooms and the lack of resources (like IT labs or teacher assistants) which would allow a more personalized attention and a way of teaching more directed to statistics itself. It is not surprising that students are demotivated and that statistics becomes one of the less popular subjects for students. It is evident that teachers –at different educational levels- should accept the fact that fast technological changes make new forms of teaching and learning foreseeable, and that we should be involved in them if we want to guide statistical education and create a real statistical culture in society.

References

1. Holmes P. (1980). Teaching Statistics – An International Journal for Teaching. Sloug: Foulsham Educational.
2. Ottaviani, M. G. (1998). Developments and perspectives in statistical education, Proceedings IASS/IAOS Joint Conference, Statistics for Economic and Social Development, Aguascalientes, Mexico, 1-4 September 1998 (CD-ROM)
3. Wild, C. y Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry (with discussion). International Statistical Review, 67(3). pp. 223-265

Statistical modeling for psychiatric nosology

Thaddeus Tarpey[1], Eva Petkova[1], R. Todd Ogden[2]
[1] New York University
[2] Columbia Univesity

Abstract
The problems of defining, diagnosing and treating psychiatric disorders are difficult and are of high-interest. This paper focuses on statistical modeling strategies for addressing this problem. In particular, the unsupervised learning methods of finite and infinite mixtures are contrasted. In precision medicine, the focus is often on treatment outcome. Can data on treatment outcomes aid in the developing neurobiological categorization approaches to differentiate psychiatric diseases?

Keywords
Aliasing; convolution; mixtures; placebo effect

1. Introduction

The term nosology refers to the branch of medical science dealing with the classification of diseases. Nosology has been a major issue in the field of psychiatric research. It is difficult to define mental illnesses such as depression and bipolar disorder etc. It is also difficult to diagnosis such medical conditions as well. Due to these difficulties, in psychiatric nosology, there is high interest in discovering biosignatures from biological measures that can be used to characterize and diagnose diseases (Insel et al., 2010). Currently, the problem of defining and diagnosing mental diseases is often based on the DSM-V, or Diagnostic and Statistics Manual (American Psychiatric Association., 2013) which is often quite problematic. For example, Ostergaard et al. (2011) show there are 1497 combinations of symptoms that can lead to a diagnosis of depression.

Highlights of the problems and issues related to psychiatric nosology are indicated by the following selected quotes from RDoC Insel et al. (2010): "Diagnostic categories based on clinical consensus fail to align with findings emerging from clinical neuroscience and genetics. The boundaries of these categories have not been predictive of treatment response... One consequence has been to slow the development of new treatments targeted to underlying pathophysiological mechanisms... the critical test is how well the new molecular and neurobiological parameters predict prognosis or treatment response... to implement1 neuroscience-based psychiatric classification."

For the diagnosis of autism spectrum disorder (ASD), the debate has centered on whether distinct disease categories exist or is there is an underlying "spectrum" of disease severity that includes for example ADHD (Grzadzinski et al., 2011). This issue was the focus of a recent article Kim et al. (2018) that proposes that ASD consists of three spectrums instead of a single spectrum. Their statistical analysis that led them to conclude ASD consists fo three spectrums was based on Latent Class Factor Analysis (LCFA). The LCFA incorporates both categorical features for presumed latent classes as well as "dimensional" features corresponding to within group continuous latent factors. Their statistical strategy in determining an appropriate model to use was to build up the model starting with only two latent classes and one continuous factor per group and then increase the number of classes and factors per class.

Another avenue of discovery in the context of psychiatric nosology is to incorporate information of defining diagnosis categories not only on symptoms but also on treatment outcome. Of course, a problem with this approach is that the type of treatment usually depends on diagnosis. However, many psychiatric illnesses are treated with the same types of medications.

A major statistical challenge for this work is the issue of aliasing. For psychiatric nosology, finite mixtures of normal distributions are an attractive approach to the unsupervised learning problem of estimating parameters for discrete disease populations. However, it has long been recognized that finite mixture distributions are often indistinguishable from homogenous continuous distributions (Pearson, 1895). If the components of the mixture distribution are normal, then the model is identifiable (e.g., Teicher, 1961). If distinct populations do not exist and disease severity varies continuously, then one can use an infinite mixture of normals which is also identifiable under certain conditions (Bruni and Koch, 1985). If it is believed there are distinct disease categories, then from a statistical point of view, this problem can be cast in the context of unsupervised learning and finite mixture model approaches may be suitable. However, if distinct disease classes do not exist, and disease severity varies continuously along a spectrum, then an infinite mixture model may be a more appropriate statistical approach to the psychiatric nosology problem. For example, Tarpey and Petkova (2010) proposed an infinite mixture model in the context of a simple regression where the predictor variable is continuous and latent. If the latent predictor is Bernoulli, then the regression model becomes a 2-component finite mixture model. Tarpey and Petkova (2010) considered a predictor variable that has a beta distribution whereby in the limiting cases, the beta can converge to a Bernoulli. In this way, the 2-component mixture and infinite mixture can be described in a single model. Tarpey and Petkova (2010) introduced this model in order to model placebo response when treating depression.

Hutson and Vexler (2018) present interesting results showing how the 4-parameter beta-normal distribution can become "aliased" with a normal distribution meaning that under particular parameter settings, the beta-normal becomes almost indistinguishable from a normal distribution. It is curious to note that strong aliasing can also occur using infinite mixtures (or convolutions) of beta and normal distributions. In our work with modeling placebo response (Tarpey and Petkova, 2010), if the population consists of placebo responders and non-responders, then an outcome variable y among drug-treated patients can be represented as a 2-component finite mixture model by

$$y = \beta_0 + \beta_1 x + , (1)$$

where β_0 is the average drug effect, β1 is the average placebo effect, and x is a Bernoulli indicator of whether the patient is a placebo responder or not, which is independent of $\sim N(0, \sigma^2)$. A realistic alternative model has the placebo response varying continuously, in which case the 0-1 Bernoulli can be replaced by a continuous latent beta variable x in (1) leading to a "latent" regression model. Various parameter configurations in this latent regression model produce distributions for y which are aliased with normal distributions. One simple illustration is to set the beta parameters to $p = q = 1\, and\, \beta_0 = 0$, $\beta_1 = 1$ and let € $\sim N(0,1)\, in\, (1)$. Then the pdf of y is $f(y) = \Phi(y) - \Phi(y - 1)$ which is essentially indistinguishable from the $N\left(\frac{1}{2}, \frac{13}{12}\right)$ distribution. Similar to the estimation problems that occur with the beta-normal distribution noted in Hutson and Vexler (2018), fitting the latent regression model with x ~ beta can lead to severe identifiability issues.

2. Methodology

The methodology we will investigate for the problem of psychiatric nosology will be called Projection Pursuit Nosology. In practice, statistical models are multidimensional as opposed to uni-dimensional. In a setting with p-dimensional measures x, the impact of a psychiatric disease may exert itself along a lower-dimensional subspace. The simplest and perhaps most useful setting is when the disruption to health generates variation in a 1-dimensional direction in the feature space.

For many variables, it may be reasonable to assume the measures vary according to a normal law for healthy individuals but if a disease is present, then that may have a skewing effect in one (or more) particular directions of the feature space. Projection pursuit (e.g., Diaconis and Freedman, 1984) is a statistical approach developed to handle such a statistical challenge.

Linear Pre-Conditioning in Clustering. Our approach will incorporate information obtained by clinician-based diagnoses and information obtained by measured features x. Specifically, we will use a clustering approach where we “learn” particular linear trans-formations of the data as a pre-conditioning before we run the clustering algorithm. This approach will lead to an iterative algorithm. The underlying idea is that if two diagnosis categories exist for a particular mental disorder, then typically there will be a strong overlap in their feature space. In order to illustrate this point, we introduce another index for cluster quality: *the variation of information* (VI) (Meilá, 2007), which is a measure of how well two clusterings of a data set coincide with each other. This measure is particularly useful in simulations where we know the true cluster memberships of data points and we can then use VI to determine how well a clustering result coincides with the another clustering. The idea here is described by the following algorithm:

0. Form an initial clustering of the features using k-means clustering.
1. Compute the VI measuring agreement between the k-means clustering and the clinician-based diagnoses.
2. Use Newton’s method to estimate an optimal direction to “stretch” (via a linear transformation) the features.
3. Re-run the clustering algorithm on this pre-conditioned data.

3. Results

The results will follow soon.

4. Discussion and Conclusion

We have proposed an unsupervised learning approach to the problem of psychiatric nosol-ogy that implements a semi-supervised clustering algorithm. The supervision comes from clinician-informed diagnosis decisions which leads to linear transformations that are then used to optimize clustering criteria.

References

1. American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Publishing, Arlington, VA, fifth edition.
2. Bruni, C. and Koch, G. (1985). Identifiability of continuous mixtures of unknown gaussian distributions. Annals of Probability 13:1341–1357.
3. Clementz, B. A., Sweeney, J. A., Hamm, J. P., Ivleva, E. I., Ethridge, L. E., Pearlson, G. D., Keshavan, M. S., and Tamminga, C. A. (2016). Identification of distinct psychosis biotypes using brain-based biomarkers. American Journal of Psychiatry 173:373–383.
4. Diaconis, P. and Freedman, D. (1984). Asymptotics of graphical projection pursuit. Annals of Statistics 12:793–815.
5. Grzadzinski, R., Martino, A. D., Brady, A., Mairena, M. A., O’Neale, M., Petkova, E., Lord, C., and Castellanos, F. X. (2011). Examining autistic

traits in children with adhd: Does the autism spectrum extend to adhd? Journal of Autism and Developmental Disorders 41:1178–1179. Doi: 10.1007/s10803-010-1135-3.

6. Hutson, A. D. and Vexler, A. (2018). A cautionary note on beta families of distributions and the aliases within. The American Statistician 72:121–129.
7. Insel, T. R., Cuthbert, B., Garvey, M. A., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., and Wang, P. (2010). Research domain criteria (rdoc): Toward a new classification framework for research on mental disorders. American Journal of Psychiatry 167:748– 751.
8. Jo, B., Findling, R. L., Wang, C.-P., Hastie, T. J., Youngstrom, E. A., Arnold, L. E., Fristade, M. A., and Horwitz, S. M. (2016). Targeted use of growth mixture modeling: a learning perspective. Statistics in Medicine 36:671–686.
9. Kim, H., Keifer, C., Rodriguez-Seijas, C., Eaton, N., Lerner, M., and Gadow, K. (2018). Quantifying the optimal structure of the autism phenotype: A comprehensive comparison of dimensional, categorical, and hybrid models. Journal of the American Academy of Child & Adolescent Psychiatry.
10. Meilá, M. (2007). Comparing clusterings: An information based distance. Journal of Mul-tivariate Analysis 98:873–895.
11. Murphy, E. A. (1964). One cause? many causes? the argument from the bimodal distribu-tion. Journal of Chronic Diseases 17:301–324.

The impact of redistribution through taxes and deductions on the income distribution – A decomposition analysis with administrative tax data from Switzerland

Oliver Hümbelin[1], Rudolf Farys[2]
[1]Bern University of Applied Sciences
[2]University of Bern

Abstract
This paper shows the potential of administrative data to grant a more complete picture of the redistributive effects of the visible (tax rates) and hidden (tax deductions) instruments of the fiscal welfare state. Based on administrative tax data from a large Swiss canton, we apply a gini-based effect decomposition to demonstrate how taxes and deductions impact the post-tax income distribution. We show that tax deductions drastically reduce the redistributive effect of taxes because lump sum deductions in a progressive tax system lead to greater tax relief for high income earners. Moreover, high income earners have additional options to claim deductions such as real-estate expenses or extra-mandatory payments to the pension scheme. In sum, deductions reduce the equalizing effect of the tax progression and therefore increase inequality.

Keywords
Redistribution effects; tax deductions; direct taxes; tax competition, Switzerland

1. Introduction

Redistributing economic resources is an important tool for the welfare state to reduce market inequalities. In this context, the OECD (2011, 2015) points out that the recent increase in disposable income inequality has been caused less by escalating markets but rather by a retreat of government. Therefore, it is important to fully understand the redistribution mechanics of a welfare state. Redistribution occurs due to social transfers and taxes. The degree of redistribution through taxes is usually attributed to the tax progression. Yet many countries also provide options for claiming deductions that alter the redistributive effect of taxes. This latter aspect is often neglected since common survey data only reports the amount of taxes paid. Administrative tax data offers a good opportunity to analyze the visible (taxes) and hidden (deductions) instruments of the welfare state in detail, as it contains complete information of the tax assessment.

The present study uses Swiss individual tax data from the canton of Aargau as an exemplar for Switzerland. Since the study is based on real tax data, it is possible to replicate the actual tax system. This permits us to evaluate every instrument of the fiscal welfare state (different taxes and deductions) with regard to its impact on income inequality. We implement a Gini coefficient based decomposition of redistributive effects by creating counterfactual "what-if" scenarios that allow us to analyze the effect of taxes if the deductions under scrutiny are included or excluded. While the main contribution of this paper is to shed light on the hidden component of the fiscal welfare state - deductions - the tax data at hand additionally provides us the opportunity to gain insight into the sometimes subtle changes of tax systems over time and how these changes affect the potential redistributive role of taxes. These insights are based on comparing the results from 2011 to those of 2001. During the intervening period, Switzerland experienced fierce tax competition between the Swiss sub-states, the cantons, which resulted in financial relief for high income earners.

2. The hidden welfare state

Deductions can be seen as an instrument of the fiscal welfare state that aims to provide tax relief to specific groups. The OECD (2010) distinguishes between distinct kinds of tax relief that have to be judged differently from a perspective of redistribution:

- *Tax exemptions*: One possibility is to exclude incomes from taxation if they fall below a certain threshold (e.g., incomes below the poverty line).
- *Privileged tax rates*: e.g. for single parents.
- *Tax credits*: e.g. loss carryforward for self-employed persons.
- *Tax deductions*: either related to certain expenses (e.g., interest costs) or standard deductions for predefined situations (e.g., child deductions).

Regarding the redistributive impact of tax deductions, one must bear in mind that their effect on post-tax income inequality is not direct but indirect. Deductions alter taxable income and tax rates, but the actual effect on the post-tax income distribution is complex and depends on the particular constellation. Theoretically, three situations can be distinguished:

- *Deductions are made equally across all income groups.* As tax rates are usually progressive, a flat deduction, however, over-proportionally favors high income filers, thus leading to an increase in inequality.
- *Deductions are more frequently used by high income filers.* Therefore, higher income taxpayers profit more and an increase in inequality is to be expected.
- *Deductions are over-proportionally used by lower income filers.* In this situation, inequality can decrease if the tax relief effect outweighs the effect of lowered tax rates.

In summary, redistributive effects are highly dependent on the specific design of the tax and deduction system as well as the pre-tax income distribution, emphasizing the need for empirical studies. To our best knowledge, there are only few studies that quantified the effect of deductions so far. Verbist & Figari (2013) showed that deductions are pro-poor in Finland, Germany, Portugal, Spain and Sweden. In Germany, for example, the deductions are aimed at pensioners who are relatively more strongly located in the lower part of the distribution. Deductions are rather pro-rich in Denmark, Ireland, the Netherlands and the UK, where either private pension contributions are deducted (Ireland, UK) or mortgage interest tax relief (Netherlands). While these studies provide first valuable insights on aggregate effects of deductions in different countries, our analysis complements these studies with detailed information on various types of deductions, as Switzerland uses a fairly complex system with numerous different deductions.

3. Data

As opposed to many other European countries where the levying of taxes is centralized, the tax system in Switzerland mirrors the historically evolved federal structure, giving a lot of power to the sub-state levels, namely the cantons and municipalities (ESTV, 2013). A total of 26 tax laws exist, with each canton having its own tax law and the municipalities and the federal state also levying taxes. This results in a multitude of direct taxes on income and wealth, which assures that each level of the state can gather the needed revenue to provide public goods autonomously.

The data we collected include all tax relevant information of the canton of Aargau and is available from 2001 to 2011. In 2011, Aargau was the fourth-largest canton of Switzerland with a population of 618’298 individuals that handed in 327’047 tax forms. With respect to economic inequality and mean income, Aargau is very close to the Swiss average. Furthermore, Aargau is also a good approximation to Switzerland with regard to its demographics and urbanity. As the argument of the present paper is of general nature and Aargau is not a special case within Switzerland, we assume our point to be valid for the whole of Switzerland as well.

4. Methodology

Definition of incomes, taxes and deductions

According to the federal structure of Switzerland, taxes are levied at three levels (federal, canton, and municipality). In addition, there is a church tax. Cantons, communities (also referred to as municipalities) and churches also levy taxes on wealth. Taxes are based on gross income, which includes all earned income, capital income and taxable social transfers. Before taxation, several different deductions can be made. Gross income minus deductions results in taxable income. Based on taxable income, either the cantonal or federal level tax rate is applied. To calculate the actual canton, municipality and church taxes, the tax determined based on the cantonal tax rate is multiplied by a factor which communities and cantons can choose for the simplified short-term management of their tax revenue to avoid the more complex legal process of adjusting the rates.

We find that more than two thirds of the tax burden is carried by canton and municipality tax, while the federal tax level has a much lower volume and the church tax is relatively minor.

For the present analyses, the deductions have been classified into six categories. *Social deductions* consist of deductions that are related to the family and health situation (second earner deduction, child deductions, health care costs, etc.). *Work-related expenses* include miscellaneous necessary costs related to employment, like the costs of commuting, weekly stays and training costs. Expenses relating to real estate like maintenance and interest costs are classified as *real estate and interest costs*. Further deductions are extra-mandatory payments to the pension scheme as well as costs of asset management and insurance costs (*Deductions related to assets and insurance)*. Finally, alimonies to partners and donations can be deducted (*Alimonies and charity*). All other deductions are classified as *other deductions.*

Decomposition of redistributive effects

Reynolds & Smolensky (1977) suggest a straightforward concept to measure the effect of redistribution by taxes. This is, as can be seen in formula (1), the difference of a Gini coefficient of pre-tax incomes (G_x) and post-tax incomes (G_{x-t}).

$$(1)\ RS = G_x - G_{x-t}$$

The present study separates the effects for single types of taxes by a sequential approach which was already used in other studies (Mahler & Jesuit, 2006; Wang & Caminada, 2011; Wang et al., 2014). Therefore, RS is depicted into three components:

$$RS_i = G_x - G_{x-t_i} = \frac{K_i * t_i}{1-t_i} - RR_i$$

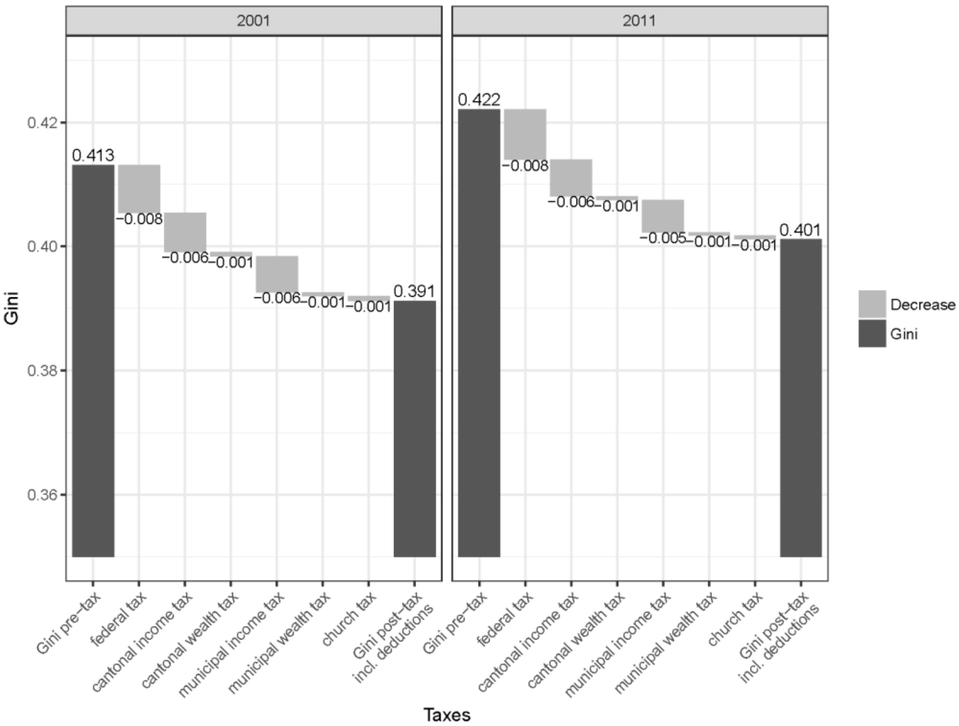
K_i is the Kakwani Index of progressivity (Kakwani, 1977) of the i-th tax, t_i is the tax rate and RR_i is the horizontal reranking effect of tax *i*. Atkinson (1979) describes this reranking as the "horizontal inequity of the tax system". As these effects might indeed be intended, horizontal effects are not discussed in the context of justice in this paper. Nonetheless, horizontal effects are calculated, as they are important in the context of the effects of deductions.

Deductions have an indirect effect on the distribution of disposable incomes as they modify the tax rate applied. To determine the effects of deductions, the tax assessment procedure of the canton of Aargau was replicated and income distributions were manufactured in a counterfactual fashion. The starting point is a distribution that would result if taxes were levied without any deductions being made beforehand. Next, the partial effect of a deduction is determined by simulating the income distribution after taxes using the i-th deduction and calculating the difference of Gini coefficients. As the first deduction usually yields the highest marginal tax relief, the effects are slightly overstated. Therefore, all effects are normalized to sum to one.

5. Results

Figure 1 shows the net effect of each of the six taxes for 2001 and 2011 and provides a visual overview of the absolute income inequality reduction effect of the single taxes. It can be seen that redistribution by income taxes is more pronounced than by wealth taxes and federal income tax has the largest effect. When comparing 2011 to 2001, inequality overall increased slightly. The effect of deductions on redistribution generally persists in 2011. One exception is the slight decrease of the impact of municipal income tax over time.

Figure 1: Partial redistributive effect of income and wealth taxes



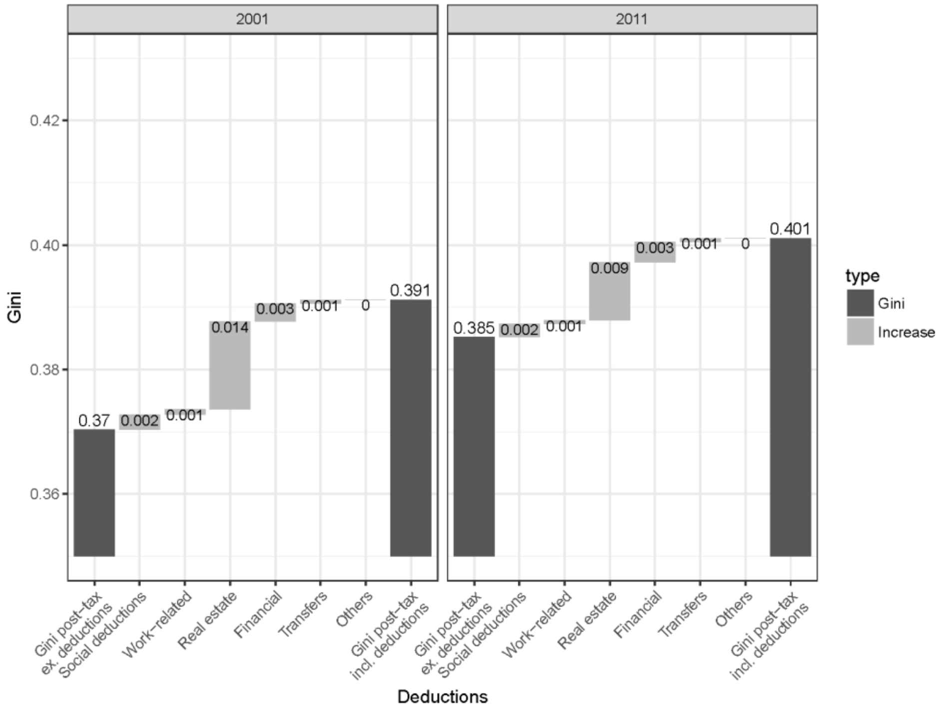
Note: Gini pre-tax refers to income inequality after social transfers but before taxes.

By comparing the absolute effects of each tax for 2001 and 2011, it can be seen that the reduced overall effect is primarily due to the reduced tax burden of the communal income tax and, to some extent, also to the cantonal income tax, while the effect of federal taxes was stable. Therefore, the proportional significance of federal taxes increased. For communal taxes, progression as well as the mean tax rate decreased. The results suggest that redistributive effects declined on the level where tax competition was most pronounced. While federal tax is exposed to international tax competition, federal units in addition compete with other cantons or even with other communities within their canton.

The figure further indicates the significance and mechanism of the different taxes. Although the federal tax makes up only 1/8 of the total tax burden, it accounts for more than one third of the total redistributive effect. According to the progressivity index, federal tax is the most progressive tax. How do deductions relate to redistributive effects?

Figure 2 provides visual access to the different redistributive effects.

Figure 2: Partial redistributive effects of deductions



The figure starts with a hypothetical value of a post-tax Gini coefficient if deductions are ignored. The deductions are then broken down into six main categories and introduced one after another so that the magnitude of the effects of each single deduction can be easily seen. Most apparent is the large effect of real-estate deductions which dominates all others in magnitude. Compared to 2001, the effects of deductions remain constant except for real-estate deduction which had a substantially lower impact in 2011. In sum, all deductions reduce the redistributive impact of taxes tremendously by -48.8% (2001) and -43.1% (2011).

Detailed analysis of each category of deductions further shows that redistributive effects vary substantially. The biggest contribution to lowering the redistributive effect results from deductions related to real estate and interest costs (e.g. mortgage interests). Comparing 2001 and 2011, it becomes apparent that this effect of real estate expenses and interest costs decreased considerably. This change is the greatest change over time and can be explained with the ongoing decrease of the mortgage reference interest rate, which was 4.25% in 2001 and 2.5% in 2011 thus leading to lower interest on debt. As a result, less interest costs had to be paid, and correspondingly less deductions were possible in 2011. Another important impact on the redistributive effect comes from deductions of costs related to assets and insurance. In particular, deductions of extra-mandatory payments to the pension scheme lower the redistributive effect. This category, moreover, gained significance over time, probably due to demographic ageing.

Deductions of work-related expenses are, in terms of volume, the second most important category. Although work-related expenses lead to an increase in progressivity, the redistributive effect of income taxes is still reduced because the tax relief this causes outweighs the higher progression. Even social deductions reduce redistribution by taxes. At the same time, social deductions caused substantial reranking effects, of 50% (2001) and 31% (2011), respectively. Therefore, social deductions are the biggest promoter of inequality between households with similar initial financial situations. Finally, it can be said that all deductions diminish the effect of redistribution via income taxes, although some increase the progressivity. This is particularly striking if the effects of work-related expenses and those related to real estate and interest costs are compared.

6. Discussion and Conclusion

While most studies focus on effects of direct taxes paid, this paper is able to expand this perspective by providing insight into the mitigating effect of the hidden part of the fiscal welfare state: deductions.

In theory, deductions can help meet social goals, e.g., by benefitting families or the ill, or by providing incentives for financially desirable behavior like saving for old age. Sometimes, deductions are also a mere hotchpotch of special interests that have accumulated over the years. In the end, the redistributive effect of deductions is determined by the degree and extent to which the different income classes actually claim deductions. The effect of deductions on redistribution is therefore hard to predict, and the present study disclosed these effects using administrative individual tax data that contains complete information on the taxing procedure from a large Swiss canton. The results indicate that deductions have a massive impact on the redistributive effect of taxes, therefore increasing inequality. For the case of the canton of Aargau, we can show that the redistributive effect of taxes was reduced by -49% in 2001 and -43% in 2011, respectively. Put simply, high income earners disproportionately profit from deductions as they are more or less flat (lump sums), while taxes are progressive, so the higher the tax paid, the higher the tax relief. Due to that mechanism, deductions increase inequality, which is not quite obvious at first glance. Second, high income earners have more options to claim deductions (e.g. related to homeownership). Another type of deduction that favors high income earners is the transfer of money to the pension system. While this is thought of as an incentive for people to save for old age, it is in practice an attractive option to lower the marginal tax rate and to flatten out incomes over the life course.

Our paper provides detailed insights of redistributive effects as part of the tax system and argues that income inequality and redistribution through taxes should be examined not only in terms of tax rates, but also with an eye on deductions in order to better understand the changing face of modern tax systems, as they can drastically moderate the direct effects of taxes.

References

1. Atkinson, A. B. (1980). Horizontal equity and the distribution of the tax burden. In H. Aaron & M. Boskin (Eds.), *The Economics of Taxation* (pp. 3–18). Washington: Brookings Institution.
2. ESTV. (2013). *Das schweizerische Steuersystem*. Bern: Eidgenössische Steuerverwaltung.
3. Kakwani, N. C. (1977). Measurement of Tax Progressivity: An International Comparison. *The Economic Journal*, *87*(345), 71–80.
4. OECD. (2011). *Divided We Stand. Why Inequality Keeps Rising*. OECD Publishing.
5. OECD. (2015). *In It Together: Why Less Inequality Benefits All*. Paris: OECD Publishing.
6. Reynolds, M., & Smolensky, E. (1977). Post-fisc distributions of income in 1950, 1961, and 1970. *Public Finance Review*, *5*(4), 419–438.
7. Verbist, G., & Figari, F. (2013). *The redistributive effect and progressivity of taxes revisited: An International Comparison across the European Union* (GINI Discussion Paper No. 88). Amsterdam: AIAS, Amsterdam Institute for Advanced Labour Studies.

Use of new technologies for the 2020 Population and Housing Census Round

Vielma, Navarro, Salcedo
University of the Philippines

Abstract
For the first time in Mexico, the 2020 Population and Housing Census data collection will be carried out through a CAPI (Computer-Assisted Personal Interviewing) scheme as the main enumeration method, but it will also the CATI (Computer-Assisted Telephone Interviewing) and CAWI (Computer-Assisted Web Interviewing) methods. These innovations, given the census structure size and the rapid changes in technologies, are a significant challenge for INEGI. Progress in census planning and field tests results will be presented, including the main challenges to be faced, the innovations considered for their implementation, as well as the successful experiences on the use of technologies for geo-referencing the information both in the data collection stage and for results dissemination.

Keywords
Population; Census; Innovations; Technologies

1. Introduction

The purpose of this document is to present the main characteristics of the 2020 Population and Housing Census; the IT strategy to be implemented, emphasizing innovations, in order to increase efficiency and control in each of its stages; as well as the main characteristics and results obtained during the field tests carried out to determine if the defined procedures are sufficiently solid to carry out this census project.

One of the most important technological innovations of this project is the census data collection through the use of mobile computing devices (MD), an area in which INEGI already has the experience of previous statistical events (surveys, economic census). Also the implementation of operating procedures supported by a set of computer tools at the service of the staff, achieving the automation of several processes.

2. Main characteristics of the Population and Housing Census 2020

The 2020 Census in Mexico, like its predecessors, will be a “de jure” census, which means that the population is enumerated in its place of usual residence. The units of observation of the census are the usual residents of the country, private dwellings and collective living quarters. The proxy respondent is the head of the dwelling or in his/her absence, a person of 18 and over who is a usual resident of the dwelling and who knows the information of its residents. The data collection is planned to take place in March 2020.

Regarding data collection forms, two population and housing questionnaires are used: a short form, consisting of around 30 questions; and a long form, designed for a probabilistic sample of around 4 million inhabited private dwellings, with approximately 75 questions, which includes the entire short form.

Additionally, the 2020 Census will provide statistics on the characteristics of the blocks' surroundings in localities of 5 thousand and more inhabitants of the country (Urban Surroundings Questionnaire), as well as the infrastructure and socioeconomic characteristics of the localities with less than 5 thousand inhabitants (Locality Questionnaire). The Social Assistance Housing Census (CAAS, for its acronym in Spanish) is also planned, which collects the characteristics of the user population, the people who work in those institutions, the buildings and the services they provide.

In order to obtain a higher response rate, a multimodal approach will be implemented, having as a main method the personal interview assisted by a MD, the self-enumeration via the Internet or the telephone assisted interview, available upon request or in case of non-response after several visits of the interviewer, as well as the use of printed questionnaires in places where the mobile device cannot be used.

3. Use of computer tools in the different stages of the Census 2020

The conceptual design contemplates the activities for the identification of information needs, which serve to determine and define the conceptual framework to which the data will be referred, the results presentation schemes, the questionnaires for its collection and the validation criteria for the revision and debugging of inconsistencies.

In order to support the conceptual design tasks, two web applications were developed, the first one called "Public Consultation to Users of the 2020 Census", which aimed to facilitate the task of identifying information needs of users, allowing them to expose systemically and simply the objectives and importance of their requirement; meanwhile, at the managerial level, it allowed to follow up on each of the requests. The second application is the "Conceptual Infrastructure of Census and Counts", whose objective is to be a repository of the conceptual frameworks for population and housing censuses, allowing the recording and updating of conceptual and operational information on the variables that have been collected in the census questionnaires, as well as reviewing changes over time.

During the operational design, the procedures and technical schemes for the data collection activities are established, as well as the administrative, organizational, control and monitoring aspects. Since the 1990 Census, INEGI has been acquiring experience in the development of systems for these tasks and for the year 2020 it is proposed to use three of them:

1. The Conformation of Areas of Responsibility System, which makes it possible to distribute equal workloads to the operational structure, automatically generating groups or segments of urban blocks or rural localities grouped by their number of dwellings and communication routes, among other factors, managing to generate areas of operational responsibility with an optimal collection order. Also, it allows staff to make manual adjustments to the conformations, applying their experience and knowledge in the field.
2. The Training System has the purpose of providing the operational personnel with the capacities and tools necessary for the execution of their functions, through automated learning modules that contain the teaching materials. This system allows giving remote advice, sending the learning activities and monitor progress in training.
3. The OPERA System, used in several stages of the information generation process, allows for timely follow-up of the progress in shaping the operational structure, systematizing the recruitment and selection of personnel, from the registration of applicants via Internet until their hiring, and follow up the logistics of the operation by controlling the materials and resources necessary for field activities and management (vehicles, computers, questionnaires, brochures, among others).

The collection of information covers the set of activities to obtain data of each unit of observation, based on the program and the established work procedures, with an operational structure and controls that promote effectiveness in each of the actions. This implies the execution of previously designed schemes for data collection, such as the preparation and distribution of support materials (cartography, manuals, instructions and catalogues), the integration of human resources, communication and consultation, under a detailed program of activities, an organizational structure and controls that should be continuously monitored from the planning phase until the closure of the census.

During this stage, the Censal Administrator application is used, in which the workloads of each operative staff are visualized and managed, allowing to reassign them if necessary. It includes a Capture Module for the timely registration of each building (Buildings List), Urban Surroundings Questionnaire, Locality Questionnaire and short and long questionnaires for the inhabited private dwellings. In this application local reports of progress and coverage can be obtained; it includes a Supervision Module in which a sampling of areas for the revision of the classification of the dwellings and inhabited condition is implemented, and through verification mechanisms it is possible to validate the quality of the information recorded by the interviewer. It also has a Cartographic Module, whose purpose is to record the cartographic updates at the block, town and road levels, allowing to digitize the polygons of new areas, cancel those that are no longer in the field and graphically represent changes such as fusions and divisions of areas, allowing immediate updating of the cartography.

On the other hand, the Module for the Coverage Verification will be implemented for the Census 2020, with the objective of facilitating the activities of the staff and, if necessary, to collect the information of the missing population in order to improve coverage in the enumeration of dwellings and people.

The Census Administrator will integrate the necessary computer tools for the execution of the Post-Enumeration Survey and a Module for the comparison between the data obtained by the enumeration and the post-enumeration.

This census also aims to promote self-enumeration, either via the internet or by telephone, so an application is available to the population in which the short form can be completed. It should be mentioned that this system will be adapted to collect the information on collective living quarters and on the staff of the Mexican Foreign Service.

The OPERA system, on the other hand, includes the module to monitor the integration of data collected in the field, allowing the generation, analysis and monitoring of certain indicators of coverage, speed and productivity in the census, important information for management decision making during the enumeration operation.

The data processing strategies include the corresponding design of the coding, validation and results generation systems with their corresponding quality controls.

In order to carry out the data processing, the Processing Monitoring and Control System will be used to monitor each of the stages. The automatic coding, by means of algorithms that recognize the textual information of the questionnaires, assigns a code of the predefined catalogues and later assigns the workloads for the computer-assisted coding. Once the coding has been completed, the automatic validation of the records is carried out. This system also includes testers of the validation criteria in order to ensure the correct application of the pre-defined algorithms. During each stage the application provides progress reports in real time. Additionally, reports on the quality of the information are generated, and a log of the changes made to the data is created. In parallel, the assignment of the geographical codes of the National Geostatistical Framework is followed up. Finally, this stage makes available the database for the activity of Figures Release, which consists of a statistical review of the historical changes of sociodemographic indicators at the state, municipal and by locality size. For this activity, the Libera System will be used, which allows the control of the revision of each indicator, but also serves as a repository for the documentation that supports changes found in some statistics.

Data processing for the publication of results is done mainly by predefined algorithms programmed ad-hoc in statistical analysis software, and in the case of products that require editing for presentation, automated processes are generated to facilitate the task, considering the editorial provisions of INEGI. There are also three web systems that allow more experienced users to access information in greater detail, these are: Consultation of Census Information System, which allows to associate census statistical information with the geographical space to which it belongs, to facilitate the interpretation of sociodemographic phenomena, through the generation of thematic maps; the Territorial Integration and Locality Consultation System, which facilitates the query of historical information of each inhabited locality of the country; and the Sociodemographic Panorama of Mexico, which is a dynamic consultation system that integrates, as a synthesis, relevant data to know the basic demographic, social and economic characteristics of the population and housing in Mexico. In addition, a series of files with the microdata of the census sample are available in different formats to facilitate its use.

4. Field tests

In the framework of the planning stage of the Census 2020, two field tests were executed in order to determine, in the operational field, the optimal characteristics of the mobile devices, the required computer tools and to define the operative procedures according to this new paradigm.

In order to test the operation of mobile devices and computer tools in different climates and areas (urban / rural) the test of operational strategy and computer equipment was conducted. During six weeks, Enumeration operations were executed, Verification and Self-enumeration, and the flow of information between each of these schemes was tested. For this test, a public bidding was carried out in which six devices of five different models were purchased, in such a way that the computer tools could be tested on equipment with various technical characteristics, all with Wi-Fi Technology, Micro USB, GPS, as well as voice and data (SIM); and computer tools for both Windows and Android were developed in order to determine the most suitable operating system for its operation.

Concerning the systems for the preparation of operational planning, recruitment and selection of staff, training, logistical and operational monitoring, integration of information to the central database, among others, the data of the observation strategy indicate quite acceptable results, providing elements that allowed to optimize its performance. Regarding the capture of the GPS coordinates of the supervisor's route, a high precision was obtained, because its visual representation is very consistent with the shape of the covered areas, with an average error of 3.8 meters. This method guarantees that the supervisor effectively made his route in the areas of his workload; this information also allows to determine the time invested and the direction of the path.

Among the main areas of opportunity, in computer science, the need was determined to implement laboratories to perform code testing and the use of best practices for the development of applications for MD with Android operating systems; the need for a tool for the remote erase or reset of the MD was also detected for the cases of robbery or loss, as well as the remote updating of the applications; in addition to the implementation of a more robust security scheme, mainly in mobile devices and in the transfer of data to the central repository.

Later, with the purpose of emulating in all the senses the activities of the Census, the Pilot Test of the Population and Housing Census 2020 was carried out, with a sample of more than 44 thousand dwellings and a duration of 20 working days. The installed offices corresponded to a municipal coordination, that is to say, one of the around 1,600 offices in which the national territory is divided for the gathering of information. All the planned operations were carried out, including the post-enumeration and the special operation for the enumeration of collective living quarters, as well as the operative verification of the inhabited condition of the buildings, which were not verified in the 2017 test.

For this operation, with the intention of testing the public bidding process to acquire the devices that will be used in the 2020 Census, a similar process was carried out to purchase the required MD. Unlike the 2017 Test, only one brand and model that complies with the minimum characteristics and technical specifications required and defined from the results of 2017 was purchased. One of the bases of this bidding considered a test of the devices, in which the transfer of information was proved through the use of a USB OTG memory, as well as the duration of at least eight hours of battery, for which tools were used ad-hoc and processes were implemented to guarantee the best device selection. The result was the purchase of 243 MD of the best offer considering the aforementioned tests.

The use of Delphi Rio de Embarcadero Rad Studio was defined for the development of the computer tools, which provides a technological platform more oriented to the development of mobile applications, in particular for devices with Android operating system, in such a way that systems were optimized for the training of the personnel, the processes of information exchange between the different operatives, the collection module of the Census Administrator, the synchronization mechanisms for the transfer of data, the security schemes from the access to the computer tools, the encryption of the information in the MD and during its transfer, until its integration in the centralized database. The modules for the generation of progress and control reports were also improved and the problems related to the use of the digital camera of the MD for the reading of QR (Quick Response) codes and the use of the GPS for the registration of the GPS coordinates were solved.

5. Technological Architecture and information security

In the Census 2020 tests, a strategy has been implemented for the transfer of information in which tools are used for the generation of packed files with security passwords, and in their interior, with the databases encrypted with the algorithm of the AES256 method (Advanced Encryption Standard with 256-bit keys). These encrypted data packages are sent through the Information Integration Module of the OPERA System, which is hosted on a secure institutional website, subsequently deposited in mass storage servers and finally they are integrated to the central database server.

The services of telecommunications, institutional network and servers of applications, databases and mass storage within the institutional network are provided by the computer infrastructure areas, who are in charge of keeping them available. The service availability scheme is configured through a load balancer of requests, through the OPERA web domain, which equally distributes each processing request between the application servers and available databases. In addition, a series of security standards and service availability tests allow planned stability throughout the whole process. For the development of Web systems, Visual Studio .NET is used, which is a complete set of tools for the generation of this type of applications; the AngularJS development framework is also used, since it provides techniques for the creation of SPA applications (Single Page Application), which benefits the performance by making requests to the application server are lighter by only requiring specific parts of the page. On the other hand, the use of JWT (JSON Web Token) implements a more secure communication and facilitates the administration of users, assigning roles and permissions. Oracle 12c is used to manage the database, providing greater security and availability of information.

6. Main technological innovations

The innovations for the realization of the 2020 Census consist on the use of mobile devices in the data collection and training; the possibility that the population can self-enumerate via the Internet or through a telephone-assisted interview; the use of QR codes; as well as the capture of the GPS coordinates during the routes of the work areas by the personnel.

The data collection of the 2020 Census, as already mentioned, will implement a multimodal approach, which contemplates the Personal Computer-Assisted Interview as the main procedure and in which phablet-type mobile devices will be used. This contributes directly to the congruence of the data and affects the coverage of areas, since it makes it possible to integrate a primary validation of the data that is recorded during the interview, to control the interview more adequately by automating the flows based on the answers of the informant, and to automatically indicate the operational situation of the work areas.

To carry out a correct identification of buildings and georeferencing of the information, the interviewer will list the dwellings and affix a label on them, whose design includes an individual unique QR code. The QR code will also be used in the self-enumeration strategy, since the interviewer will leave invitation letters in the dwellings with no response or upon request of the informant. These invitations contain a QR code that includes a unique username and password for each dwelling, in addition to the necessary instructions for the registration of information via the Internet and an assistance telephone number. The use of these codes will also strengthen the post-enumeration operation.

The use of GPS has several purposes, among which can be listed: to indicate the proximity of the staff in the field, not only to the work areas but also to the northwest corner where the reconnaissance route and / or the information gathering should start; to record the coordinates of the personnel paths in the work areas, mainly the supervisor and the interviewer, which facilitates the work of supervision, verification and post-enumeration, since these coordinates can be represented in the geographic module for access to the work areas of the staff responsible for these operations; georeferencing each of the buildings and dwellings of the country, which together with the QR code of the labels will allow updating the National Inventory of Dwellings and will be the start of the statistical registration of the real estate in the country.

7. Discussion and Conclusion

The use of mobile devices undoubtedly offer advantages over the use of paper. The adoption of new operating procedures and information processing, which start even before the data collection itself with the digital cartography update, requires special attention and care.

Regarding the applications, it is necessary to accelerate the life cycle of the development in order to minimize the maintenance stage during the operation, and to revise the flow of the information. This implies the realization of a greater number of tests, both in controlled conditions and in the field, which range from field work to the production of information in order to reduce the time for the presentation of results.

Additionally, the country's security conditions require us to prepare plans for the mitigation of risks in the field, such as paper-assisted interviews, so the traditional processing methods must be available for these cases, even if the scale is much smaller. Therefore, in the planning process, it is necessary to consider multimodal schemes for data collection and for the integration of the data bases coming from the devices, paper questionnaires and self-enumeration by Internet. All this efforts will allow to improve the consistency of the information, to publish results timely and will generate savings mainly by eliminating the data capture staff.

References

1. Instituto Nacional de Estadística y Geografía (INEGI) (2011). Síntesis metodológica y conceptual del Censo de Población y Vivienda 2010. Aguascalientes, México, INEGI.
2. Instituto Nacional de Estadística y Geografía (INEGI) (2015). Encuesta Intercensal 2015: síntesis metodológica y conceptual. Aguascalientes, México, INEGI.
3. Instituto Nacional de Estadística y Geografía (INEGI) (2017). Aspectos Generales del Censo de Población y Vivienda 2020. Consulta Pública 2017. Aguascalientes, México, INEGI. (Internal document).

Multiple co-clustering with heterogenous marginal distributions and its application to identify subtypes of depressive disorder

Tomoki Tokuda[1,2], Junichiro Yoshimoto[2], Yu Shimizu[1], Kenji Doya[1]
[1]Okinawa Institute of Science and Technology Graduate University, JAPAN
[2] Nara Institute of Science and Technology, JAPAN

Abstract

With the advent of sophisticated data acquisition methods, huge amounts of data have become available. Cluster analysis is a powerful data mining tool to reveal the underlying heterogeneous structure of objects in data. Recently, co-clustering method gains much attention for its attempt to reveal relationships between object and feature, hence capturing a possible interplay between them. However, in a big dataset, multiple cluster structures may exist, where cluster solutions differ depending on the features that one focusses on. Furthermore, the marginal distribution of feature that characterizes a cluster may be heterogeneous, e.g., Gaussian, Poisson or multinomial. To cope with these challenges in big data analysis, we developed a novel multiple co-clustering method. Our method is based on nonparametric Bayesian mixture models in which features are optimally partitioned for each cluster solution. This feature partition works as feature selection for a particular cluster solution, screening out irrelevant features. For mixture components, we assume Gaussian, Poisson or multinomial distributions (pre-specified, but the mixing of these different types in data is allowed). We present the theoretical foundation of our method, and show how our method works on real data. The demonstration data is based on our recent study on identification of subtypes of depressive disorder using high-dimensional data of different modalities such as functional Magnetic Resonance Imaging (fMRI), clinical questionnaire scores, and genetic polymorphism.

Keywords

Multi-view clustering; Mixture models; Feature selection; MRI

1. Introduction

We consider a clustering problem for a data matrix that consists of objects (or subjects) in rows and features (variables, or attributes) in columns. Clustering objects based on the data matrix is a basic data mining approach, which groups objects with similar patterns of distribution. As an extension of conventional clustering, a co-clustering model has been proposed, which captures not only object cluster structure, but also feature cluster structure (Lazzeroni & Owen, 2002; Gu & Zhou, 2009; Madeira & Oliveira, 2004). In the present paper, we focus on a specific type of co-clustering, so called 'check board' where both objects and features are exclusively partitioned (Fig.1A; features are partitioned based on their distribution patterns). Yet, the co-clustering method (as well as conventional clustering methods) does not always work well for real data, because real data may have different 'views' that characterize multiple clustering structures (Fig. 1B; Muller et al., 2012; Niu et al., 2010).

To find the underlying multiple clustering structures, however, determination of the number of views is not straightforward. A promising approach is based on nonparametric mixture models assuming multivariate Gaussian mixtures for each view proposed by Guan (2010). In this approach, the full Gaussian model for covariance matrices is considered, and the numbers of views and of object clusters are inferred in a data-driven way via the Dirichlet process. Such a method is quite useful to discover possible multiple cluster structures by screening out irrelevant features, when these numbers are not known in advance. However, this method suffers from the drawback that features need to belong to the same distribution family, which severely limits its application, because real data often include both numerical and categorical features. Further, its application is rather limited to low dimensional cases ($p < n$), because in high-dimensional cases, sample size to infer the full covariance matrix of Gaussian distribution may be insufficient, resulting in overfitting.

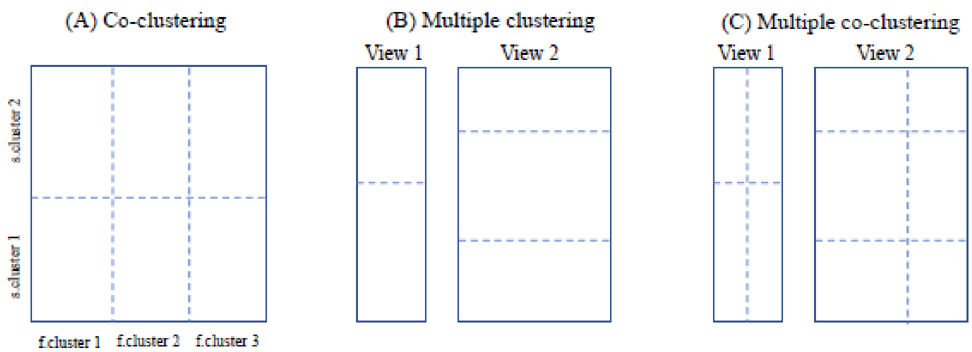


Figure 1. Illustration of clustering structures. In each panel, the horizontal axis denotes feature index, while the vertical axis subject index. Subjects and features are sorted in the order of their cluster memberships. Dashed lines denote boundaries between subject clusters or between feature clusters. Note that in Panels (B) and (C), subjects members in view 1 and view 2 are the same, but they are differently sorted following their cluster memberships in each view.

To address the aforementioned problems, we consider a multiple clustering framework in which we can make the best use of co-clustering structure that is not prone to overfitting. We propose a novel multiple clustering method (referred to hereafter as the multiple co-clustering method) based on the following extension of the co-clustering model. First, we consider multiple views of co-clustering structure (Fig.1C), where a univariate distribution is fitted to each cluster block (Shan & Banerjee, 2008). Second, for each cluster block, the proposed method simultaneously deals with an ensemble of several types of distribution families such as Gaussian, multinomial and Poisson distribution. Obviously, the first extension enables our model to fit high-dimensional data, while the second enables it to fit data that include different types of features (numerical, categorical, and integer).

2. Methodology

Our method is based on a Bayesian approach, which models view/co-clustering structures and instances in each cluster block. We outline relevant parameters and an inference method in the following sections (see more details of this section in Tokuda et al., 2017).

View and co-clustering structure: We denote a $n \times d$ data matrix as $\boldsymbol{X}$ with n subjects (or, objects) and d features. To infer the multiple co-clustering structures as seen in Fig.1C, we introduce latent variables of labelling feature and subject memberships. First, both for view and co-clustering structures, we introduce a $G \times V$ latent matrix $\boldsymbol{Y}_j$ for the feature j, where V and G are the number of views and feature clusters, respectively. In this notation, a view membership and a feature cluster membership are defined as 0 (false) or 1 (true). For instance, $\boldsymbol{Y}_j$= ((0, 0, 0, 0)$^{\mathrm{T}}$, (0, 0, 1, 0)$^{\mathrm{T}}$, (0, 0, 0, 0)$^{\mathrm{T}}$) denotes that the feature j belongs to view 2 and feature cluster 3 in that view (the superscript T denotes the transpose of vector; each vector denotes a feature cluster membership for a particular view). As the definition of membership implies, only one element in $\boldsymbol{Y}_j$ is 1, while the remainder of them 0.

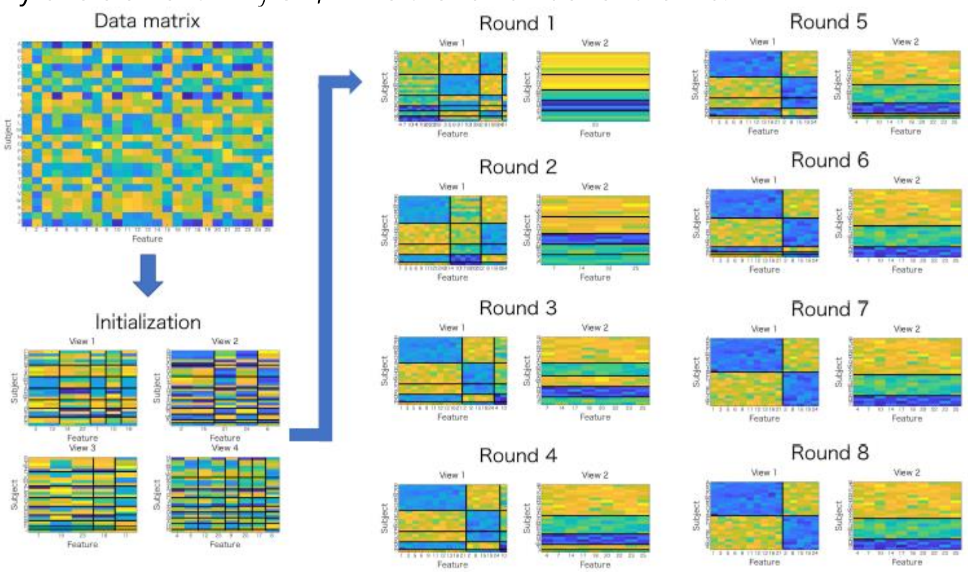


Figure 2. Demonstration of the multiple co-clustering algorithm. The algorithm starts with a random initialization for an arbitrary number of views. Subsequently, hyperparameters in parameter distributions are iteratively updated. The resultant clustering structure in each round of updating is shown in a panel. It is observed that the algorithm has converged at Round 7. The horizontal axis denotes feature indices, while the vertical axis subject indices.

Note that the number of feature clusters may differ among views, but to avoid cluttering, we take G as the maximum number of feature clusters over views, allowing for empty clusters. Similarly, for subject cluster memberships, we introduce a $K \times V$ latent matrix **Z**i, which denotes subject cluster memberships for the subject i in views. For instance, **Z**i=((0, 1, 0, 0)^T, (1, 0, 0, 0)^T, (0, 0, 0, 1)^T) denotes that the subject i belongs to the subject cluster 2 in view1, the subject cluster 1 in view 2, and subject cluster 4 in view 3. Hereafter, we denote a cluster block for feature cluster g and subject cluster k in view v, as cluster block (g, k, v).

Probabilistic model in a cluster block: For numerical features, we assume that instances in each cluster block follow a univariate normal distribution with specific mean and variances. We denote a pair of mean and variance for cluster block (g, k, v) by $\boldsymbol{\theta}_{g,k,v.}$Using this notation, the logarithm of conditional likelihood of **X** is given by

$$\log p(X|Y,Z,\Theta) = \sum_{v,g,k,i,j} \mathbb{I}\left(Y_{j,g,v} = 1\right)\mathbb{I}\left(Z_{i,k,v} = 1\right)\log p\left(X_{i,j}|\theta_{g,k,v}\right) \quad (1)$$

where $\mathbb{I}(x)$ is an indicator function, i.e., returning 1 if x is true, and 0 otherwise; **Y**={**Y**$_j$}; **Z**={**Z**i}; $\boldsymbol{\theta} = \{\boldsymbol{\theta}_{g,k,v}\}$; $Y_{j,g,v}$ an element (g, v) of $\boldsymbol{Y}_j$; $Z_{i,k,v}$ an element (k, v) of **Z**i. If we know the true values of **Y**, **Z** and $\boldsymbol{\Theta}$, Eq.(1) simply becomes a sum of the logarithm of density function, which is evaluated in each cluster block for instances that belong to the corresponding cluster block. Similarly, for categorical and integer features, we define multinomial (including a categorical distribution as a special case) and Poisson distributions, respectively. The conditional log-likelihood of the concatenated data of these different types of distributions is simply a sum of the right hand side in Eq.(1) for each type of distribution.

Prior and posterior for parameters: Our objective is to infer unknown parameters **Y**, **Z** and $\boldsymbol{\Theta}$ from the data matrix **X**. We infer these in a Bayesian approach, introducing prior distributions for these parameters. Since we do not know the true number of views and the true numbers of feature clusters and subject clusters, we assume infinite numbers for views and clusters, which is modeled in the framework of Dirichlet process. For probabilistic parameters $\boldsymbol{\Theta}$, we assume conjugate priors for each type of distribution. Using a joint prior distribution $p(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Theta})$ for these parameters and the likelihood $p(\boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Theta})$ in Eq.(1), the posterior is given by

$$p(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Theta}|\boldsymbol{X}) \propto p(\boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Theta})p(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Theta}).$$

Inferences: The MAP (Maximum *a posteriori*) estimates $\boldsymbol{Y}$, $\boldsymbol{Z}$, and $\boldsymbol{\Theta}$, which maximize the left-hand side in Eq.(2), provide an optimal solution for estimation of multiple co-clustering structures and probabilistic density distributions in each cluster block. However, direct maximization of $p(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Theta}|\boldsymbol{X})$ is rather difficult because of the complexity of relationships among the parameters. Hence, we approximate it by means of variational inference (Bishop, 2006), denoting an approximate density function as $q(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Theta})$, in which we assume independence among these parameters. In this framework, we aim to find the best approximation in terms of Kullback-Leibler (KL) divergence between $q(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Theta})$ and $p(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Theta}|\boldsymbol{X})$, which reduces the problem into optimization of hyperparameters in these distributions. The algorithm of optimization proceeds as follows. First, the data matrix is randomly partitioned into views in which subject- and feature- cluster memberships are randomly allocated to subjects and features. From this initial configuration, the hyperparameters are iteratively updated to monotonously decrease the KL divergence. When the first round of updating is completed, the next round of updating is subsequently carried out, which is repeated a number of rounds until the KL divergence converges. A visual demonstration using an artificial data is shown in Fig.2.

3. Results

As an application to real data, we present the results of analysis on a dataset of MDD (Major Depressive Disorder) in our recent study (Tokuda et al., 2018). The objective of the study is to identify subtypes of MDD in a data driven manner.

Background: It is well known that MDD is heterogenous in such characteristics as clinical presentation, progression, treatment response, genetics, and neurobiology. This heterogeneity hampers progress in identifying the cause of MDD and its effective treatment. To overcome this problem, several studies have been so far conducted to identify subtypes of MDD in a data-driven manner, relying on clinical questionnaires. However, the results of these studies either conflict or they simply identify clusters related to depression severity, which does not provide conclusive evidence for subtypes of depressive symptoms. Furthermore, these studies are based on clinical questionnaires without considering of biological substrates. We address this question by means of the multiple co-clustering method. We apply this method to a combination of several dataset modalities, such as functional connectivity (FC) data in resting state fMRI (i.e., a subject stays at rest while fMRI images are scanned), clinical questionnaires, and gene expression data.

Data: The number of subjects is 134, comprising of 67 MDD and 67 healthy subjects. Our data comprise several modalities, such as FC data in resting state fMRI, clinical questionnaires, and biological data, which are mixed up with 2800 numerical, 100 categorical, and 2 integer features. For simultaneous analysis of all datasets, we concatenated them into a single data matrix (subjects are common to the sub-datasets). For FC data, functional connectivity between regions of interest (ROIs) in the brain were evaluated, which is defined as correlation coefficient of average time series of Blood Oxygen Level Dependent (BOLD) fMRI signals between ROIs. Since ROIs consist of 78 brain regions, this resulted in 2701 features (78 × 77/2) for FC. Clinical questionnaire data comprise several types of scores that measure depression severity (e.g., Beck Depression Inventory, BDI), personality (e.g., NEO personality Inventory), and life experiences such as Child Abuse Trauma Scale (CATS). In particular, BDI and Hamilton Rating Scale of Depression (HRSD) are important indicators for measuring the severity of depression focusing on its several symptoms. Hence, we also included question items about these indicators as a categorical feature. The same clinical questionnaires were administered to both depressive and control subjects. Importantly, for depressive subjects, these were administered before the onset of Selective Serotonin-Reuptake Inhibitor (SSRI) treatment. Moreover, some indices, such as BDI and HRSD, were again administered six weeks and six months after the treatment (these features are denoted with endings of $6w$ and $6m$). We expected that inclusion of repeated measurements in these scores would provide useful information on treatment effect. Beside functional connectivity factors, we included several biological elements: sex, age, Brain Derived Neurotrophic Factor (BDNF), Cortisol, Single Nucleotide Polymorphisms (SNPs), and DNA Methylation.

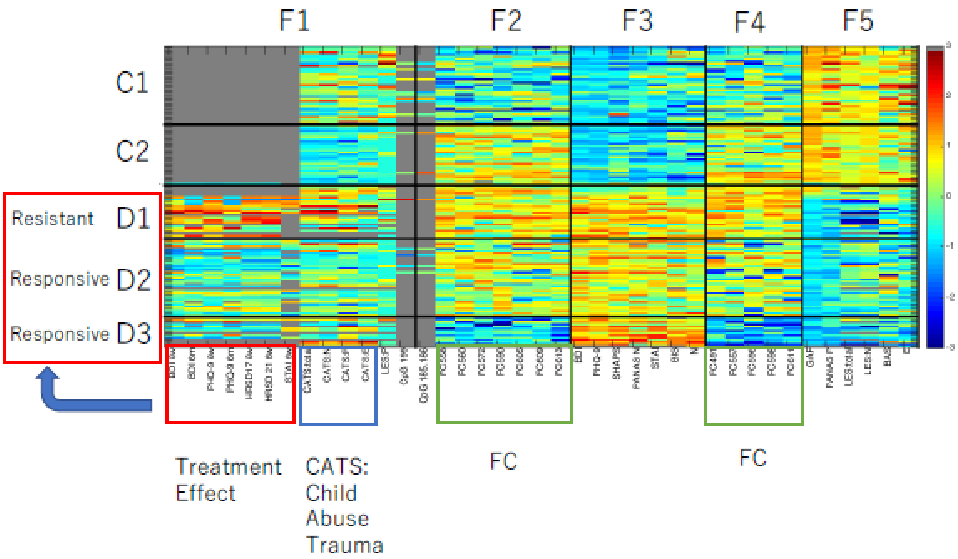


Figure 3: Heatmap of view 10. MDD subjects and healthy subjects are clearly separated: Subject cluster C1-2 consist of only healthy subjects, while D1-3 consist of only depression subjects. Further, feature cluster F1 includes features related to treatment effect, which characterize D1-D3 as treatment resistant, responsive and responsive.

Pre-processing of data:
For the numerical features, we standardized each feature using its mean and standard deviation, based on all subjects (ignoring missing entries). In contrast, for categorical and integer type of features, we did not carry out pre processing. Note that we left missing entries as they were, because our clustering method can handle them in a Bayesian manner without explicit substitutions.

Results of multiple co-clustering: The multiple co-clustering method yielded 15 views in which the view index is based on the number of features in a view in the descending order. Specifically, we analyzed view 10 more in detail (Fig.3). This view consists of five subject clusters, which match the label of control/depression well: Two clusters for control subjects (subject clusters C1, C2 in Fig.3) and three clusters for depressive subjects (subject clusters D1, D2, D3). View 10 contained several non-FC features that discriminate well among subject clusters. Further, the view had a high proportion of depression-related features (60%), including 39 numerical features and 19 categorical features, but none of the integer features. Since the categorical features do not clearly distinguish between subject clusters, we focus only on numerical features. These numerical features are clustered into five feature clusters F1-5. It is noteworthy that the view contains a feature cluster (F1) that is related to the after-treatment status of depression, namely, the features in a red box labelled 'Treatment Effect' in Fig.3. A higher value in these features implies that a patient is not remitted. Therefore, we can characterize D1, D2, and D3 as resistant, responsive, and responsive clusters for SSRI treatment. Furthermore, features related to CATS are included in the same feature cluster F1. These results suggest that the subject clusters D1, D2 and D3 for depressive subjects may be related to after-treatment status of depression, which might be further related to stress experiences during childhood. Lastly, feature clusters F2 and F4 are related to specific functional connectivity in fMRI image data (i.e., angular-gyrus related FC), which suggests a possible association between the subject clusters and neural substrates.

This result of cluster analysis raises the possibility of prediction of treatment outcome prior to SSRI treatment. In this regard, we explore several important implications drawn from the result. First, subject clusters can be represented by a small number of relevant features. Since in our clustering method each feature cluster consists of similar (i.e., highly correlated) features, a feature cluster can be represented by a reduced number of these features. It turns out that the subject clusters D1, D2 and D3 are represented by CATS scores (associated to feature cluster F1), and the first principal scores of angulargyrus related FC (associated to F2 and F4; we simply refer to it 'AG related FC score'). These features can indeed explain the resultant subject cluster membership properly (Fig.4A). In the scatter plot of Fig.4A, AG-related FC scores discriminate between subjects in D3 and other subjects. On the other hand, CATS scores do not discriminate a single class by itself, but they discriminate between subjects in D1 and D2, once subjects in D3 are sorted out. This observation motivated us to consider a classifier that consists of the following steps. First, we classify subjects into either D3 or non-D3 based on AGrelated FC scores. A subject with low scores in AG-related FC is classified into D3, otherwise into non-D3. Subsequently, the non-D3 subjects are classified into either D1 or D2 based on CATS scores: A subject with low scores in CATS is classified into D2, otherwise into D1. This procedure of classification is summarized in Fig.4B. Since these subject clusters correspond to degrees of remission of SSRI treatment as well, this classifier leads to predictions of whether SSRI treatment may be effective, prior to the onset of treatment. We can interpret this classification as follows. For subjects in D2 and D3, SSRI treatment may be appropriate (low after-six-week BDI scores), while it may not for those in D1.

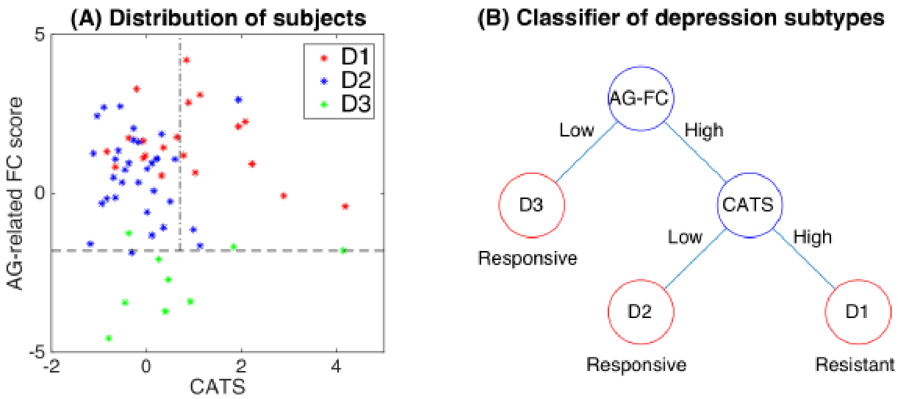


Figure 4. Classification of D1, D2 and D3. Panel (A): Distribution of subjects. Panel (B): A classification tree based on Panel (A).

4. Conclusion

We have developed a novel multiple co-clustering method based on nonparametric Bayesian mixture models. This method can reveal the underlying multiple view structures in which co-clustering structures with different types of distributions are imbedded. Applying this method to a concatenated dataset of different modalities of depression data, we have identified a relevant view for subtypes of depression. Further analysis of this view implies a possible classification of treatment-resistant depression and non-treatment-resistant depression prior to the onset of SSRI treatment. This result is useful for further investigation of classification of MDD patients from the perspective of treatment effect.

References

1. Bishop, C.M. (2006). Pattern Recognition and Machine Learning. Springer.
2. Gu, Q. & Zhou, J. (2009). Co-clustering on manifolds. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 359–368.
3. Guan, Y., Dy, J.G., Niu, D. & Ghahramani, Z. (2010). Variational inference for nonparametric multiple clustering. In: MultiClust Workshop KDD-2010.
4. Lazzeroni, L. & Owen, A. (2002). Plaid models for gene expression data. Statistica Sinica, **12**, 61–86.
5. Madeira, S.C. & Oliveira, A.L. (2004). Biclustering algorithms for biological data analysis: A survey. IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), **1**, 24–45.

6. Muller, E., Gunnemann. S., Farber. I. & Seidl, T. (2012). Discovering multiple clustering solutions: Grouping objects in different views of the data. In: Data Engineering (ICDE), 2012 IEEE 28th International Conference, 1207–1210.
7. Niu, D., Dy, J.G. & Jordan, M.I. (2010). Multiple non-redundant spectral clustering views. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), 831–838.
8. Shan, H. & Banerjee, A. (2008). Bayesian co-clustering. In: Data Mining, 2008. ICDM'08, Eighth IEEE International Conference, 530–539.
9. Tokuda, T., Yoshimoto, J., Shimizu, Y., Okada, G., Takamura, M., Okamoto, Y., Yamawaki, S. & Doya, K. (2017). Multiple co-clustering based on nonparametric mixture models with heterogeneous marginal distributions. PloS ONE, **12**, 0186566.
10. Tokuda, T., Yoshimoto, J., Shimizu, Y., Okada, G., Takamura, M., Okamoto, Y., Yamawaki, S. & Doya, K. (2018). Identification of depression subtypes and relevant brain regions using a data-driven approach. Scientific Reports, **8**, 14082.

Rank-one matrix factorization for co-clustering: An optimal transport perspective

Charlotte Laclau[1], Franck Iutzeler[2], Ievgen Redko[1]
[1]Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d Optique Graduate School, Laboratoire Hubert Curien UMR 5516 F-42023, Saint-Etienne, France
[2] Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK 38000 Grenoble, France

Abstract
In this paper, we first present a novel approach for the co-clustering of matrices rows/columns using entropy-regularized optimal transport. We show how this technique is connected to several seemingly unrelated unsupervised learning algorithms including matrix factorization. Using this observation as a starting point, we developed two novel algorithmic solutions for the co-clustering problem based on OT and rank-one matrix factorization. We believe that our work provides a new point of view for several unsupervised learning techniques that helps to gain a deeper understanding about the general mechanisms of co-clustering.

Keywords
Co-clustering; Optimal Transport; Matrix Factorization

1. Introduction

In many real-world applications, data is often collected in form of observations related to joint interactions between two types of entities. Usually, these interactions are represented as a matrix whose modes correspond to the underlying entities and whose entries reflect the type or the intensity of the interactions between them. Co-clustering exploits the correlation between the rows and the columns of such matrices and allows to identify local structures given by groups of rows which share similar patterns for a specific subsets of columns. These methods are often categorized into two main frameworks, notably probabilistic and metric-based methods. Despite a large variety of existing approaches, several studies proved that sometimes seemingly distinct methods actually optimize the same objective function. This is the case, for instance, for k-means and EM for Gaussian mixture models clustering where the former can be shown to be a special case of the latter. On the other hand, different versions of non-negative matrix factorization minimize an objective function similar to a constrained k-means problem (Ding et al, 2010) and thus are intrinsically linked with probabilistic models. All these connections are of great importance for scientific community as they provide fundamental understanding of mechanisms behind unsupervised learning in general.

In this paper, we study the connection between entropy-regularized optimal transport and rank-one matrix factorization and show how they allow to obtain the co-clustering partitions with an automatic detection of the number of rows and columns clusters. Our interest for these methods is due to the emergence of optimal transport in machine learning on one hand and by the longstanding presence of matrix factorization on the other. We would like to underline that our paper studies and presents the connection between the above mentioned methods in an exploratory fashion: our foremost goal is to illustrate the intuition and provide a new point of view allowing to understand the link that exists between these two learning approaches, that seem completely different at first sight. The rest of this paper is organized as follows. In Section 2, we present the preliminary knowledge related to optimal transport problem and its entropic regularized version. In Section 3, we explore the relationship between regularized optimal transport and matrix factorization for the co-clustering problem. Finally, we conclude our paper and give several challenging future research directions that can be derived from it.

2. Optimal Transportation of Discrete Measures

Optimal transport considers the problem of finding a mapping transporting one probability measure to another one in a way that minimizes the transportation cost. Hereafter, we consider the formulation of Kantorovich (Kantorovich, 1942) for discrete measure, that consists in finding a coupling matrix $\gamma \in \mathbb{R}^{m\times n}$ such that $\gamma_{i,j}$ is defined as the fraction of mass transported from the $i-th$ bin of the source distribution α (i.e. xi) to the $j-th$ bin of the target distribution $\beta(i.e.\ y_i)$, belonging to the simplex of size$-m$ and size$-n$, respectively. The set of admissible couplings writes:

$$\Pi(\alpha,\beta) \coloneqq \{\gamma \in \mathbb{R}_+^{m\times n} :\ \gamma\mathbf{1}_n = \mathbf{a}, \gamma^\top 1_m = \mathbf{b}\},$$

where $\mathbf{1}_n$ is the size$-n$ vector of ones. To find an optimal coupling among the admissible ones, one has to define a cost matrix $\mathbf{M} \in \mathbb{R}_+^{m\times n}$ that models the cost of moving source bin i to target bin j represented in our case by x_i and y_i, respectively. For instance, if $d_{\mathrm{x}} = d_y$[1] a natural choice for $\mathbf{M}$ is the squared Euclidean distance, i.e. $M_{ij} = \left\|\mathrm{x}_i - \mathrm{y}_j\right\|_2^2$. In the following, and for complexity reasons, we focus on the proposition of Cuturi (2013) that consists in adding a regularization to promote couplings that are close to the trivial coupling $ab^\top \in\ \Pi(\alpha,\beta)$ in the sense of the Kullback-Liebler divergence. In this case, optimal transportation problem writes:

[1] Note that when this is not the case, one may use the Gromov-Wasserstein distance defined for metric-measure spaces of different size. We refer the reader to Peyre et al (2016) for further details.

$$\min_{\gamma \in \Pi(\alpha,\beta)} \langle \mathrm{M}, \gamma \rangle_F - \lambda E(\gamma), \qquad (1)$$

where λ ≥ 0 is a regularization parameter. This regularization allows to obtain smoother and more numerically stable solutions compared to the original case. Indeed, Sinkhorn's theorem Sinkhorn and Knopp (1967) tells us that for λ > 0, (1) has a unique solution that can be obtained by left and right scaling of the Gibbs kernel $e^{-M/\lambda}$ (where the exponential is taken element-wise) to the prescribed sums of the admissible couplings:

$$\gamma_\lambda^\star(\alpha, \beta) \coloneqq \arg\min_{\gamma \in \Pi(\alpha,\beta)} \langle \mathbf{M}, \gamma \rangle_F - \lambda E(\gamma) = \mathrm{diag}(\mathbf{u}) e^{-M/\lambda} \, \mathrm{diag}(\mathbf{v}),$$

where **u** and **v** are two non-negative scaling vectors uniquely defined up to a multiplicative factor, that can be efficiently computed using Sinkhorn's algorithm.

3. Our Contributions

3.1 Co-clustering through Optimal Transport

Let us denote by X and Y two random variables taking values in the sets $\{x_i\}_{i=1}^m$ and $\{y\}_{i=1}^n$, where x_i, y_i correspond to i^{th}row (instance) and column (feature), respectively. We further assume that the joint probability distribution between X and Y denoted by $p(X,Y)$ is estimated from the data matrix **A**. The problem of co-clustering consists in jointly grouping the set of features and the set of instances into homogeneous blocks by finding two assignment functions C_r and C_c that map as follows: $C_r : \{ x_1, \dots, x_m\} \rightarrow \{\hat{x}_1, \dots, \hat{x}_g\}, C_c : \{y_1, \dots, \hat{y}_k\}$ where g and k denote the number of row and columns clusters, and discrete random variables $\hat{X}$ and $\hat{Y}$ represent the partitions induced by X and Y, i.e., $\hat{X} = C_r(X) =$ and $\hat{Y} = C_C(Y)$.

The main idea is to consider the problem of finding a coupling between the rows and columns of a given data matrix **A**, represented by a uniform sum of m and n Diracs defined over a set of points in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Formally, these rows and columns empirical measures can be defined as:

$$\boldsymbol{k_r} := \sum_{i=1}^{m} \delta_{X_i} / m \text{ and } \boldsymbol{k_c} := \sum_{i=1}^{n} \delta_{X_i} / n.$$

For a given cost matrix **M** where $M_{i,j}$ stands for the distance between line i and column j, the authors obtain an optimal coupling $\gamma_\lambda^\star(\boldsymbol{k_r}, \boldsymbol{k_c})$ using Sinkhorn given by:

$\gamma_\lambda^\star(\boldsymbol{k_r}, \boldsymbol{k_c}) = \mathrm{diag}(\mathbf{u}) e^{-M/\lambda} \, \mathrm{diag}(\mathbf{v})$.

The obtained solution can be seen as a joint probability distribution implying that the scaling vectors **u** and **v** can be seen as estimated rows and columns marginal distributions. For more details, see (Laclau et al, 2017).

3.2 From Optimal Transport to Matrix Factorization

Now, let us further consider the objective function related to the optimal transportation problem with Kullback-Leibler regularization presented in above and defined as:

$$\mathcal{L}(\gamma; M; \alpha, \beta) = \langle M, \gamma \rangle_F + \lambda KL(\gamma \| \mathbf{ab}^\top).$$

The minimization of $\mathcal{L}$ is usually performed w.r.t. the coupling γ with a and b being fixed vectors from size-m and size-n simplexes, respectively. However, for the co-clustering problem introduced above, we are interested in finding the scaling vectors **u** and **v** for a given data matrix **A** directly and thus propose to consider the following minimization problem:

$$(u, v) \coloneqq \arg\min_{(u,v)} \mathcal{L}\,(\mathbf{A}; \mathbf{M}; \mathbf{u}, \mathbf{v}),$$

where contrary to the original problem the minimization is performed over **u** and **v**. We note that this problem simply amounts to the minimization of $\mathrm{KL}(\gamma \| \mathbf{ab}^\top)$ with $\gamma = \mathbf{A}$ that does not necessarily represent some joint distribution. Instead, this problem aims to find two vectors **u** and **v** that have sufficient entropy or small enough mutual information w.r.t. the matrix of interactions **A**. In the general case of a matrix **A** $\in \mathbb{R}^{m \times n}$ with potentially missing values, the calculation of **u** and **v** amounts to solving the following matrix factorization problem:

$$\min_{(\boldsymbol{u},\boldsymbol{v}) \in \mathbb{R}^m \times \mathbb{R}^n} \sum_{\substack{i,j=1 \\ \boldsymbol{A}_{i,j}\ exists}}^{m,n} \boldsymbol{A}_{i,j} \log\left(\frac{\boldsymbol{A}_{i,j}}{u_i v_i}\right) - \boldsymbol{A}_{i,j} + u_i v_i.$$

The connection between optimal transport and matrix factorization can also be shown using a different point of view. When α and β are defined as uniformly weighted sums of Diracs, the Kullback-Leibler divergence term in (3) becomes:

$$\mathrm{KL}(\gamma \| \mathbf{ab}^\top) = \mathrm{KL}(\gamma \| \mathbf{11}^\top / mn).$$

One may further note that as λ grows, the solution of the regularized optimal transport becomes closer to the uniform distribution and thus its factorization becomes:

$$\mathbf{11}^\top / mn \underset{\lambda \to \infty}{\sim} \gamma_\lambda^\star(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathrm{diag}(\mathbf{u}) e^{-\mathrm{M}/\lambda}\, \mathrm{diag}(\mathbf{v})$$
$$(1/\mathbf{u})\ (1/\mathbf{v})^T \simeq e^{-\mathrm{M}/\lambda} mn.$$

The obtained equation can be equivalently seen as a rank-one matrix factorization of the Gibbs kernel $e^{-\mathrm{M}/\lambda}$ that depicts the joint interactions between two sets of objects underlying α and β. On the other hand, in the extreme case when $\lambda \rightarrow \infty$, the Gibbs kernel $e^{-\mathrm{M}/\lambda}$ becomes a matrix of ones leading to a degenerate solution where **u** and **v** are vectors of ones.

Finally, we note that when λ does not tend to infinity, one may retrieve a particular type of tri-factor factorization from the solution of the regularized OT problem. To see that, we can rewrite the factorization of the optimal coupling matrix as follows:

$$e^{-\mathrm{M}/\lambda} = \mathrm{diag}\left(\frac{1}{\mathbf{u}}\right)\gamma_\lambda^\star(\boldsymbol{\alpha}, \boldsymbol{\beta})\mathrm{diag}\left(\frac{1}{\mathbf{v}}\right),$$

where vectors **u**, **v** and the optimal coupling matrix $\gamma_\lambda^\star(\boldsymbol{\alpha}, \boldsymbol{\beta})$ are unknown. This formulation is reminiscent of a popular co-clustering method called Tri-NMF (Ding et al, 2008) formulated as the following optimization problem over $\mathbf{W} \in \mathbb{R}_+^{m\times k}, \mathbf{S} \in \mathbb{R}_+^{k\times g}, \mathbf{H} \in \mathbb{R}_+^{g\times n}$ for some $\mathbf{A} \in \mathbb{R}_+^{m\times n}$:

$$\mathbf{A} \simeq \mathbf{WSH}, \mathbf{W}^T\mathbf{W} = \mathbf{I}, \mathbf{H}^T\mathbf{H} = \mathbf{I},$$

where **W** and **H** are low-rank approximations of features and instances, respectively, **S** is a matrix of associations between them and the orthogonality constraints are added to ensure the uniqueness of the factorization and the ease of interpretability for clustering. From this formulation, we note that co-clustering solutions provided by **u** and **v** can be seen as full-rank approximations of matrices **W** and **H** while the meaning behind $\gamma_\lambda^\star(\boldsymbol{\alpha}, \boldsymbol{\beta})$remains similar to that of **S**. Furthermore, matrices $\mathrm{diag}\left(\frac{1}{\mathbf{u}}\right)$ and $\mathrm{diag}\left(\frac{1}{\mathbf{v}}\right)$ corresponding to **W** and **H** naturally fulfill the orthogonality constraints up to a scaling factor. This point of view tells us that regularized optimal transport can be equivalently seen as both rank-one and full-rank matrix factorizations of the Gibbs kernel $e^{-\mathrm{M}/\lambda}$ where in both cases the latent low-rank structure for rows and columns should be extracted from one-dimensional objects **u** and **v**. Figure 1 shows the cluster-generated vectors obtained by them on the same synthetic data set. The simulated matrix contains 60 variables and 60 instances, and presents a block structure with 3 homogeneous groups, or clusters, for both instances and variables. From this figure, we can see that all cluster-generated vectors present similar changes in slope, at close but slightly different locations, revealing the same type of steps for both **u** and **v.**

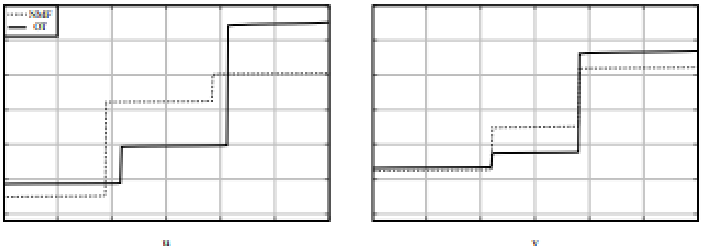


Fig. 1: Comparison of the vectors obtained with Matrix Factorization (NMF), and optimal transport (OT) on a simulated squared data matrix with a 3 × 3 block structure.

3.3 Detecting the number of clusters

In order to map coordinates of the obtained cluster-generating vectors to final clustering partitions, we concentrate our attention on a popular methods based on the Potts problem (Weinmann et al, 2015). In general, for a given size$-n$ vector $\mathbf{u}$, this method aims at finding a piece-wise constant vector x by solving an optimization problem composed of two terms: (1)a fitting term taken as the $\ell_p$norm $\|\mathbf{x}-\mathbf{u}\|_p^p$ for some $p \geq 1$; (2) a regularization term, controlled by an hyper-parameter $\lambda_{\text{reg}} > 0$, that penalizes $D \in \mathbb{R}^{n \times n-1}$ such that $[D\mathrm{x}]_{i=1} = (\mathrm{x}_{i+1} - \mathrm{x}_i)$ for $i = 1, \ldots, n-1$. In order to properly solve this problem, we propose to use the ℓ_p *-sorted-Potts* problem which penalizes the increments of the *sorted* vector. Denoting by $\mathfrak{s}$ the sortin operation, the ℓ_p -sorted-Potts problem writes:

$$x^{\star} = \arg\min_{x \in \mathbb{R}^n} \|\mathbf{x}-\mathbf{u}\|_p^p + \lambda_{reg}\|D\mathfrak{s}(\mathrm{x})\|_0.$$

We propose to apply this regularisation to the vectors resulting from both OT and MF.

4. Conclusion

In this paper, we studied regularized optimal transport and its application in co-clustering. More precisely, we showed that this former can be expressed as rank-one matrix factorization problem that produces cluster-generating vectors of a given data matrix. Finally, we proposed a regularization strategy that enhances the block structure of the obtained cluster-generating vectors. This work can be extended in several original research directions. In the near future we would like to exploit a long-known equivalence between optimal transport and k-means Caas and Rosasco (2012) to extend this study further to other families of learning methods.

References

1. Caas GD, Rosasco L (2012) Learning probability measures with respect to optimal transport metrics. In: NIPS, pp 2501–2509
2. Cuturi M (2013) Sinkhorn distances: Lightspeed computation of optimal transport. In: NIPS, pp 2292–2300
3. Ding C, Li T, Peng W (2008) On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing. Comput Stat Data Anal 52(8):3913–3927
4. Ding CHQ, Li T, Jordan MI (2010) Convex and semi-nonnegative matrix factorizations. IEEE Trans Pattern Anal Mach Intell 32(1):45–55
5. Kantorovich L (1942) On the translocation of masses. In: C.R. (Doklady) Acad. Sci. URSS(N.S.), vol 37(10), pp 199–201
6. Laclau C, Redko I, Matei B, Bennani Y, Brault V (2017) Co-clustering through optimal transport. In: ICML, pp 1955–1964
7. Peyre G, Cuturi M, Solomon J (2016) Gromov-wasserstein averaging of kernel and distance ´ matrices. In: ICML, pp 2664–2672
8. Sinkhorn R, Knopp P (1967) Concerning nonnegative matrices and doubly stochastic matrices. Pacific Journal of Mathematics 21:343–348
9. Weinmann A, Storath M, Demaret L (2015) The l 1 -potts functional for robust jump-sparse reconstruction. SIAM Journal on Numerical Analysis 53(1):644–673

Developing effective models of public-private partnerships (PPPs) in National Statistical Ecosystems: Lessons from Asia

Matthew Shearing
GOPA Statistics and independent consultant on official statistics

Abstract
The data revolution and adoption of the Sustainable Development Goals (SDGs) gave a new impetus for the development of national statistical systems (NSSs). These developments call for more support to the NSSs and finding efficient methods for the private and public sectors to work together to share knowledge and deliver high quality statistics. To ensure effective use of the limited financial support available, traditional donor-recipient models may need to be revised and new means of collaboration developed that sustain innovation and maintain quality in official statistics. The Asian experience is rich in terms of learning from how a wide variety of organisations have worked together in different ways as providers of data, innovation, and of skills and knowledge. This paper aims to analyse examples of experiences in Asia in PPPs inform the development of PPPs globally.

Keywords
official statistics; PPPs; SDGs; financing; capacity development

1. Introduction

The Paris Declaration on Aid Effectiveness and the Accra Agenda for Action underline the need for "building partnerships for development. Such partnerships are most effective when they fully harness the energy, skills and experience of all development actors"[1]. Moreover, the financial context for official statistics is highly challenging. The OECD's Development Cooperation Report 2017 identifies an annual funding gap of $685 million for SDG data in developing countries[2]. However, in 2016, the Global Partnership for Sustainable Development Data estimated that only $240 million would be delivered for IDA-eligible countries[3].

In terms of meeting these gaps, current models of public-public sector cooperation and PPPs are clearly not meeting demand. However, the global statistical system has recently focussed its attention on the private sector as a provider of data rather than a provider of wider services. But is this missing a

[1] http://www.oecd.org/development/effectiveness/34428351.pdf
[2] http://www.oecd.org/dac/development-co-operation-report-20747721.htm
[3] https://opendatawatch.com/wp-content/uploads/2016/09/development-data-funding-2016.pdf

major opportunity? While public-public sector cooperation will inevitably remain important in managing and delivering improvements in statistical capacity, there is significant evidence that the private sector can be further leveraged to increase statistical capacity, increase the funding available for it, lower costs, and improve impact. There is an urgent case therefore that action should be taken to explore the best ways of utilising this potential.

2. Methodology

This paper draws on lessons from Asia in published material, discussions with development actors in the public (NSOs and international organisations) and private sectors. It includes reflection on the author's personal experiences and discussions with a wide range of other independent consultants with in-depth experience in statistical capacity-building in Asia, with a range of private sector companies, and working in and with NSOs and international statistical organisations[4]. The conclusions and results are provisional to stimulate to debate at the ISI World Statistics Congress in 2019 and to promote the unearthing of wider evidence to inform the development of effective models of PPPs.

3. Results

3.1 The traditional role of PPPs in developing official statistics

Typically, the private sector has worked with the public sector on a range of projects from institutional reforms to support for innovation and the use of new data sources and so on. Donors/ beneficiaries may hire either a company or an individual from the private sector to deliver these services. These services are often delivered through a combination of different types of actors across the public and private sectors. In many cases, a private sector company will be selected to manage overall delivery by a consortium of public and private sector actors. In other cases, donor organisations, often due to prohibitive bureaucracy in contracting services via the market-place, will sub-contract individuals and directly manage or carry-out project implementation themselves. This is often the case in the Asian Development Bank and United Nations. In some cases, private sector companies are employed directly by governments. This may become an increasingly important model in terms of filling urgent skills gaps, such as in data science, particularly given the private sector's natural strength as innovators and ability to rapidly mobilise skills in response to emerging needs.

[4] *The author in particular wishes to thank Misha Belkindas, Mike Hughes, Jean-Michel Durr, and Richard Roberts*

3.2 The strengths of the public, private, and other sectors

Individual technical experts (private sector)

These typically have experience working in a specific topic area(s) in an NSO(s) and/or international organisation(s) and in working on capacity-building projects in various countries. They bring in-depth technical expertise, particularly in the day-to-day and long-term challenges of producing official statistics, good (and bad) practices across the globe, working in and tailoring inputs to individual countries and specific regional or national NSS contexts.

Advanced NSOs (public sector)

These bring in-depth and up-to-date knowledge of advanced techniques and challenges in NSOs and can form very effective working relationships with fellow professionals. Where NSOs are experienced in capacity-building projects, such as Statistics Netherlands, they can also provide some of the benefits of independent technical experts. However, the availability of their experts during the project is often limited due to the demands of their domestic responsibilities, naturally high costs, and their expertise may be focussed on technical solutions that are inappropriate for certain developing countries. Project management, including financially transparent accounting systems, are not a core strength. However, a limited amount of advanced NSOs have also developed in-house project management capability in capacity-building. But there are challenges which arise when and because these operate outside of a competitive market-place (see section 3.3).

Technical service providers specialised in official statistics (private sector)

These organisations are the most common actors to lead tenders for and implement commercial projects. With the right incentives and contractual frameworks these organisations can run large and small-scale capacity development projects efficiently. They bring expertise in combining successful project management with a contextual understanding of the needs of official statistics, often including internal expertise. They are particularly strong in understanding the practicalities of project implementation, mobilising the right subject matter experts, improving cost-effectiveness, developing effective work plans and responding to challenges, robust and transparent financial management systems, and coordinating partners and stakeholders across sectors.

Figure 1: Example of NSOs working with the private sector on capacity development

Many PPPs include an advanced NSO (or NSOs) working under the coordination of a technical service provider from the private sector to the benefit of a less advanced NSO. The Lao Statistics Bureau (LSB) has worked with at least three different advanced NSOs under this arrangement. The LSB Deputy Director General set out at the World Data Forum (WDF) in Dubai 2018 that such "Twinning" under the coordination of private sector had relieved LSB of the administrative burden of recruiting, managing experts, and monitoring their interventions. It was also noted that this kind of PPP had allowed for:

- access to a wide pool of highly qualified experts
- ability to finance early mobilisation and rapid deployment
- efficient coordination of team members' work
- flexibility in meeting changing priorities
- potential to ensure the sustainability of the project's impact
- enhanced strategic planning, communication and coordination with development partners

Statistics Netherlands, having worked as part of such arrangements in Laos also pointed out at the WDF that working with a technical service provider from the private sector was important given their experience in managing complex projects, financial administration, and bringing together a mix of suppliers, including different NSOs with different skills, independent experts, and other private sector parties, such as IT experts and data innovators.

Technically-specific organisations (private/other sectors)

Organisations from many sectors can contribute to projects in specific technical area(s), based on their area of expertise, such as in Data Science, IT systems, or statistical methodology. These include academic institutions with training units/capability in technical areas or regional training institutes. This category also includes subject-specific private companies who are often highly innovative, aware of broader perspectives and techniques across different sectors, and may have experience in contributing to capacity development projects. These may be most valuable where their expertise is outside of the expertise typically held by statisticians, such as IT systems, data science, or communications. Training or other services from these organisations is most effective where they have experience in contributing to capacity development projects.

Figure 2: Example of a PPP in the use of mobile phone data for official statistics in Indonesia

Badan Pusat Statistik (BPS), the Indonesian NSO, recently signed an MoU with a private sector company, Positium. This sets the framework for the private sector's provision of support to BPS in data science techniques, based on experience applying this knowledge in other NSSs, in terms of developing official tourism statistics using mobile phone data from Indonesian telecoms giant Telkomsel. The private sector company and BPS worked together in securing financing and approval from the Indonesian Ministry of National Development Planning (BAPPENAS) and in getting the right mobile phone data from Telkomsel.

Less advanced NSOs (public sector)

Notwithstanding the overall benefits and desirability of 'South-South' development, less advanced NSOs have a lot to offer each other statistically. But the commitment of resources is potentially even more challenging than with advanced NSOs. A private sector partnership may be a helpful way of creating the necessary funding and an overall efficient framework for cooperation. Moreover, NSOs with lower levels of development often benefit from learning from those NSOs who have recently developed their standards from a similar base level. This type of involvement may be particularly effective where mixed with a highly advanced NSO; ie 'triangular' cooperation and learning.

National and regional actors (public/other sectors)

While NSS bodies are the most important actors in implementing change in their organisations, all NSSs operate in a political environment. It is thus necessary to identify the key government and public institutions that require consultation on project design and engagement during implementation. This may need to be formalised, such as through a Memorandum of Understanding. This may include sub-national governments or regional institutions (such as ASEAN, ADB). There may also be independent national or regional institutions that can sometimes be formally part of a partnership or could be usefully consulted on project design and implementation – such as national statistical societies. These can usefully contribute to long-term sustainability, building trust with users and provide independent and well-pointed advice.

3.3 The Economics of PPPs in statistical capacity development

It seems that some NSOs, international organisations, and donors may not have adequately considered the full economic costs and value-for-money of their choices in respect of the models and partners selected for capacity development. There has been a tendency to work directly with fellow public sector professionals. This is potentially driven by convenience (avoiding prohibitive bureaucracy), habit, a lack of trust in other sectors, legal concerns, and mis-understandings about how the marketplace can optimise value-for-money. It is necessary to be clear about how the market-place, or lack of one, can affect the financial requirements and efficiency of expenditure on developing statistical capacity.

Price mechanisms

The public sector is highly constrained in the efficiency in which it can deliver services due to the pricing mechanisms involved in its work. Many NSOs can be reluctant to release their staff from their daily activities to support activity on the development of other NSOs, particularly their most knowledgeable staff. Where they can, naturally they will charge high prices. Existing NSO staff, while valuable to the beneficiary NSO, will therefore come at a high price. Even where there is a private sector company leading a partnership, there is limited room for negotiation and driving down the overall cost. For public bodies such as NSOs and international statistical organisations, there is no market incentive to decrease staff pricing which is often fixed by internal regulation. This factor can inflate the costs of capacity development projects delivered by the public sector and minimise the availability of appropriate expertise.

A leading role for private sector technical service providers can minimise these risks. This includes their ability to mobilise independent expertise at market prices and find the best value among possible contributing NSOs. However, it may also be beneficial to consider ways in which pressures on NSO prices can be reduced and overall positive outcomes increased. We may therefore consider how contributing NSOs can be attracted to a capacity-building project not just as a source of income but because it will also lead to strategic internal gains eg: through the development of its staff, improving corporate knowledge including innovations which they can apply in their own organisation.

Competition mechanisms

While allowing private sector providers to compete on price can lead to significant reductions in costs, particularly in driving down the costs of human resources, the market also provides incentives for them to compete to provide the highest quality services, including in ensuring that outcomes are sustainable. Experience shows that this has often been a successful mechanism in terms of private sector companies innovating and investing resources (both financial and intellectual) in creating proposals for the best way to deliver statistical objectives. Project tenders are typically scored and awarded based on overall price, methodology, and the team structure which is proposed; the latter being informed by the strengths of the different types of partners proposed, as set out in section 3.2.

Contracting mechanisms

Following successful competition between private sector service providers, ongoing incentives need to be refined to ensure that the private sector follows through on its potential to achieve high performance and sustainable improvements. Managed correctly, this can provide greater assurance on the quality of outputs and the efficiency of inputs than the often ambiguously worded non-financial Memorandums of Understanding which underpin typical bilateral NSO-NSO cooperation on capacitybuilding.

However, under-performance against objectives has been a problem in some statistical capacitybuilding projects involving PPPs, as it has been for much public-public sector cooperation. This has resulted from complex factors that are difficult to assign to individual actors. The structures of contracts are potentially to blame, as they have perhaps relied too heavily on negative incentives such as the possibility of contract cancellation and/or non-payment which are very difficult to enact. More importantly there has been a lack of positive incentives for implementers to solve emerging challenges. Improvements could include quantitative and qualitative objectives that could be rewarded with additional financial gains if exceeded. Interventions using the private sector could also better integrate sustainability by building in targets/incentives for beneficiaries and implementers that are linked to positive developments after a project's formal closure; a 5-10-year timeframe after project completion may be appropriate.

Investment seeking

Private sector actors are naturally willing to invest in securing future revenue. This factor can lead to mutually beneficial outcomes and the leveraging of finance from potential donor institutions. A key question then is about how this potential for the private sector to raise revenue is currently being suppressed. Evidence of any globally significant amount of private sector fund-raising is thin.

A major challenge is probably the willingness of public sector institutions in statistics to support this process. PPPs could be very powerful in lobbying potential donors for financial support for statistical/data improvements, particularly when such improvements are presented as adding value to wider investments by the donor (such as their spending on the wider development of a country in question). This lack of public sector willingness goes beyond the national level; international developments in statistics seem too exclusive of the private sector in the coordination of the response to capacity development needs. Trust building activities and the exchange of information between the sectors has the potential to release the confidence required by the private sector to invest in financeraising activities.

Furthermore, many private sector technical service providers in official statistics are sizeable enough to provide pre-financing of projects, which represents a temporary investment allowing for capacity developments to take place in a timely fashion.

Figure 3: Example of the potential of PPPs to raise funds for statistical development

In 2017, GOPA Statistics, a private sector technical service provider, successfully lobbied the Government of Luxembourg to invest in capacity development work in Laos. This involved a partnership between the Luxembourg NSO (STATEC) and GOPA in both the lobbying and delivery of the project which has allowed for the continuation of important structural statistical reforms in Laos.

Financial transparency

The need for more transparency in the financing of development - improving the quantity and quality of financing data - is a widely recognised requirement[5]. Where capacity development is market-driven this is a fundamental benefit given the transparency required in budgeting and expenditure. This contrasts with the hidden costs of public sector activities, such as additional staffing costs (eg business class flights) which then limits the budgets available to NSOs to provide on-the-ground support to other NSOs. Moreover, the overall transparency of the private sector's motivations contrasts sharply with the more political elements of the international interaction between public sector bodies. These motivations may be 'politically efficient' but they are rarely transparent and economically efficient, thus jeopardising the efficient distribution of available funds for capacity development.

3.4 Wider barriers to effective PPPs

The public sector in general seems to remain overly nervous about working with the private sector. Recent efforts by the public sector to reach out to the private sector to solve the Agenda 2030 challenge have been largely

[5] *For example: http://www.oecd.org/dac/development-co-operation-report-20747721.htm*

driven by the necessity to access new data sources, and then only with significant caution. In terms of solving the funding crisis for statistical development, the engagement seems to have been excessively sparse.

This caution is to some extent legitimately driven by both legal and philosophical concerns. In the case of gaining access to new data sources, it is important however to reflect on the fact that these concerns are gradually being eroded given the will to address them. Therefore, this could also be the case in increasing the use of the private sector in wider capacity development. In any case, it is now increasingly evident that it is an existential matter for official statistics to evolve to be a more outward-focussed multi-sectoral enterprise. But trust remains a major barrier. There are still strong perceptions that private sector is solely motivated by making profits and not the public interest, and moreover that it is morally wrong to spend public money on private sector involvement in official statistics. These perceptions should be challenged[6].

A less psychological barrier holding back this potential are the bureaucratic structures in which some public sector organisations operate. Some NSOs for instance are not able (by regulation) to operate in a commercial venture. Some international organisations are highly constrained by the resources and time required to organise the procurement of a commercial tender. However, regulations and bureaucracies are not static. They can evolve in line with internal and external political priorities. So perhaps the fundamental barrier is a lack of senior leadership in making the necessary changes to free up the potential of the private sector.

4. Discussion and Conclusion

This paper has made the case that the global statistical community must pay more attention to the use and development of PPPs in capacity development. Analysis of the economics of statistical capacity development seems to indicate that an expanded role for the private sector in statistical capacity development could release significant benefits. However, further discussion and evidence gathering is encouraged to establish:

- What further lessons can we learn from experiences of PPPs in capacity development projects to improve their design and the use of the private sector more broadly?
- What can be done to reduce the barriers in NSOs and international organisations to release the potential of PPPs?
- What can the private sector do to enhance its suitability for greater involvement in capacity development?

[6] https://undataforum.org/WorldDataForum/suspicious-minds-trust-and-releasing-the-potential-of-publicprivate-partnerships-in-statistical-capacity-development/

Quality control in large-scale onine knowledge integration using OBJECTIVE

Rense Lange[1,2]
[1]Global Psytech, Selangor, Malaysia (rense.lange@globalpsytech.com)
[2]Laboratory for Statistics and Computation, ISLA, Vila Nova de Gaia, Portugal

Abstract
As is useful in essay-grading, product-evaluation, and some crowd-sourcing tasks, Multi-Faceted Rasch Scaling (MFRS) models provide simultaneous diagnostic information concerning products' (essays) perceived quality, buyers' (test-takers) overall preferences or trait-levels, raters' leniency / severity, as well as rating scale usage. Unfortunately, the traditional approaches to fitting MFRS models are suitable for batch processing only, thus greatly limiting their real-time (online) potential usage. However, it proved possible to structure the PAIRS parameter estimation method such that most of the estimation computations are already performed during data entry. Simulations indicate that the resulting system is highly efficient and for samples with up to simulated 2,000,000 cases in an offline context, parameter updates took less than 0.001 seconds. The findings were sufficiently encouraging to form the basis for the OBJECTIVE system that can update model parameters nearly instantaneously, thereby creating actionable quality-control information for use in actual practice.

Keywords
OBJECTIVE; Multi-Faceted Rasch Scaling; Grading; Rater severity; Rater bias

1. Introduction

Modern psychometrics provides powerful practical tools for test construction, quality control – including identifying and removing item and test biases [see, e.g., 1] – and test equating (i.e., deriving equivalent scores using different items). Psychometrics aims to assess respondents' (test-takers) traits or abilities based on their answers to a series of questions (items). Its methods are widely applied in education and psychology and usage is increasing in other online areas like product evaluation [2].

Here I focus on Multi-Faceted Rasch Scaling (*MFRS*), as was pioneered by Linacre [3]. *MFRS* can include factors beyond test-takers and items – the classical application being essay grading where *MFRS* is used to quantify the role played by the essay graders. In addition to providing parameter estimates, it also serves a quality control function by providing extensive model fit information. The central feature of essay grading (i.e., evaluation by proxy) are shared by applications such as crowd-sourcing, product rating, and online service reviews [2]. Because essay writing and evaluation procedures are familiar to nearly everyone, I use essay grading terminology throughout the following.

There already exists excellent software to estimate *MFRS* models and to assess model fit [4, 5]. However, this software is *batch-oriented,* for use when data collection is completed. Yet, in practice it is often desirable to identify and correct problems much earlier. For instance, in a project that required timely grading of some 150,000 student essays [6], *MFRS* was used to identify poorly performing graders. However, model fitting could not be completed before graders already had started their next working day and this greatly limited its' usefulness. Similar concerns play a role in other contexts [2].

This paper grew out of the *OBJECTIVE* project that aims to fit *MFRS* models online in "real-time" - at least in an essay grading context, while providing useful quality control information. The current emphasis is the parameter estimation aspect via the *PAIRS* approach [7], and issues related to quality control and model fit are outside the scope of the current paper. The basics are discussed in some detail [7, 8] and this is followed by a presentation of the results of various computer simulations that are aimed at discovering the approach's strength and weaknesses for applications like *OBJECTIVE*.

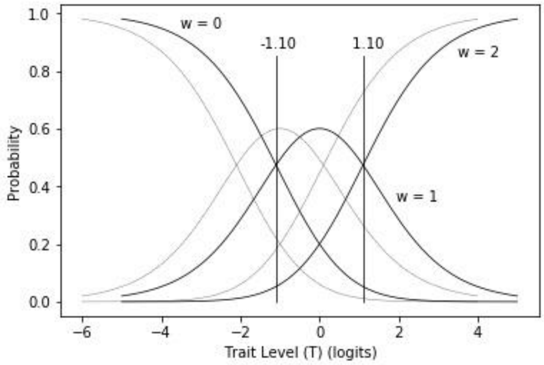
2. Methodology

Essay grading is characterized by four sets of the parameters [2]: (1) D_i, the difficulty of the item, question or problem i; (2) T_j, the ability, capability, or trait level of the person answering the questions; and (3) S_k, the severity of the rater / evaluator who assigns an (ordinal) grade to evaluate the answer or product. Raters' judgments typically take the form of a rating scale, i.e., raters choose one option out of a set of ordered response categories. Accordingly, a fourth set of "step" parameters $\{F_w\}$ is used to represent the points at which ratings $w = i$ and $w = i-1$ occur with equally probability. Together, these parameters define the following probabilistic response model (see, e.g., [3]):

$$log(P_{ijkw} / P_{ijk(w-1)}) = T_j - D_i - S_k + F_w, \quad (1)$$

where P_{ijkw} represents the probability that we observe a rating equal to w rather than (w-1) with $w = 0, 1, \ldots$. Notice that T, D, S, and F are all expressed in the same metric, i.e., the log-odds (or *logits*) as defined by the left side of Equation 1.

Figure 1: Probability of Observing Ratings w= 0, 1, and 2 as Respondents' Trait Levels (T) Increase.



It is possible to solve for the various parameters (see e.g., [3]), which yields the equations (not shown here) that describe the probability of each answer. For instance, Figure 1 (solid lines) describes P_{ijkw} for a three-category rating scale with step values $F_1 = -1.1$ and $F_2 = 1.1$, $S = D = 0$, with T varying along the X-axis. Note that as T increases the lowest rating (i.e., $w=0$) becomes less likely than the value $w=1$, which is then gradually superseded by the value w=*2*. Note that $w=0$ and $w=1$ are equally likely to occur at $T = F_1 = -1.1$ and that $w=1$ and *2* occur with equal probability at $T = F_2 = 1.1$.

The parameters combine additively in Equation 1, thereby facilitating interpretation. For instance, if we had used a rater whose severity is 1 logit *less* than the current rater (i.e., $S = -1$) then the probabilities are described by the dotted lines in Figure 1, i.e., the curves shift to the left by 1 logit.

Estimation. The estimation of *MFRS* parameters traditionally relies on a Joint Maximum Likelihood Estimation approach (*JMLE*) [2]. While *JMLE* is satisfactory in a batch-oriented context, it is not suitable for an incremental approach where model parameter estimates may be needed while data is being gathered. In *JMLE* all parameter estimates are mutually dependent during estimation, as updating the trait parameters T requires D, S, and F, updating D requires T, S, and F, etc. Thus, each update necessarily involves several passes through the entire dataset and over 200 iterations may be required to reach convergence [4]. Accordingly, computational demands are typically dominated by T (respondents' traits or abilities) as this is almost always the facet with the greatest number of levels.

***PAIRS* for Binary Items**. To avoid *JMLE's* computational demands we instead use the *PAIRS* approach first proposed by Rasch [7], see [8]. To be sure, *PAIRS'* computational demands are *also* dominated by the levels of T. However, with some pre-processing many required computations can already be performed during data gathering while new data are entered. With all pre-processing completed, obtaining the final parameter estimates requires very limited additional processing. Importantly, the final estimation step *in no way* depends on the number of cases already processed.

PAIRS can be illustrated via the behaviour of the simplest case, i.e., binary (e.g., True / False) Rasch items (without using any raters), as would apply to a standard multiple-choice type test. Here,

$$log(P_{ij} / (1 - P_{ij})) = T_j - D_i, \qquad (2)$$

where P_{ij} is the probability that person j will correctly answer item i. Solving for P_{ij} in Equation 2 yields:

$$P_{ij} = exp(T_j - D_i) / (1 + exp(T_j - D_i)). \qquad (3)$$

Let a_{vi} denote the outcome of person v taking some item i, i.e., 0 when incorrect and 1 when correct. Consider $a_{vi} = 1$ given that the outcomes for two items i and j sum to unity, i.e., $a_{vi} + a_{vj} = 1$:

$$P\left(a_{vi} = 1 \middle| a_{vi} + a_{vj} = 1\right) = \frac{P(a_{vi} = 1)P(a_{vj} = 0)}{P\left(a_{vi} = 1 \,\&\, a_{vj} = 0\right) + P(a_{vi} = 0 \,\&\, a_{vj} = 1)}$$
$$= \frac{e^{D_j}}{e^{D_i} + e^{D_j}} \qquad (4)$$

If we denote the number of test-takers who answer i correctly and j incorrectly by b_{ij}, then the probability (4) can also be estimated empirically via:

$$\frac{b_{ij}}{b_{ij} + b_{ji}} = \frac{e^{D_j}}{e^{D_i} + e^{D_j}},$$

Dividing by b_{ji} yields:

$$\frac{b_{ij}/b_{ji}}{b_{ij}/b_{ji} + 1} = \frac{e^{D_j - D_i}}{e^{D_j - D_i} + 1}, \text{ and hence } \frac{b_{ij}}{b_{ji}} = e^{D_j - D_i}.$$

Provided that both b_{ij} and b_{ji} are non-zero, taking logarithms of both sides yields:

$$\log\left(\frac{b_{ij}}{b_{ji}}\right) = D_j - D_i\ . \qquad (5)$$

In other words, much of the information needed to compute the D_i can be kept updated by maintaining the appropriate b_{ij} counts in a matrix. As will be shown below, obtaining the actual D values involves division and taking logarithms as in Equation 5 to obtain the matrix *log(R)*. Then, averaging the rows of *log(R)* yields least-square estimates of the D_i [8]. Note that no assumptions are made concerning the distributions of the model parameters, which supports the approach's robustness.

Figure 2: PAIRS algorithm applied to 2500 simulated items and 25 simulated persons (see text).

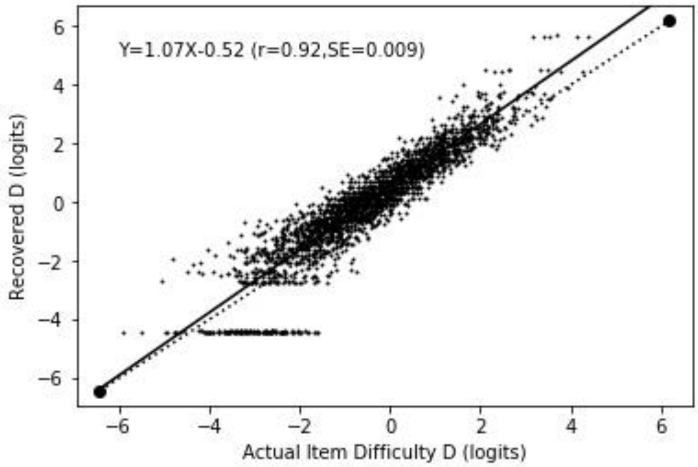


Illustration. Simulations indicate that *PAIRS* works surprisingly well for binary items. For instance, 2500 item difficulty parameters D_i were drawn from a normal distribution with M = *-0.5* and SD = *1.5* (i.e., *N(-0.5, 1.5)*). Also, twenty-five person trait level parameters T_j were obtained drawing from *N(0, 1).* Equation 5 was then applied to the resulting 25 x 2500 matrix of 0 and 1s to recover the D_i. It can be seen in Figure 2 above that this unusually small number of respondents sufficed to recover the 2500 D_i with considerable accuracy ($r = 0.92$, $SE = 0.009$) – albeit with poor results for extreme values of D. To bring SE down to below 0.001 required just 50 simulated persons ($r = 0.98$).

Figure 3: Frequency (F) tables of item x ratings pairs (left) and rater raw sums (right).

		Item 1		Item 2		Item 3		
		0	1	0	1	0	1	
Item 1	1						1	D_{11}
	2							D_{12}
Item 2	1							D_{21}
	2		1				1	D_{22}
Item 3	1							D_{31}
	2							D_{32}

	Rater 1	Rater 2	Rater 3	Rater 4	
	x-1	x-1	x-1	x-1	
Rater 1 x	0	13	5	7	S_1
Rater 2 x	3	0	3	2	S_2
Rater 3 x	15	3	0	2	S_3
Rater 4 x	8	14	0	0	S_4

Raters and Rating Scales. Equation 1 allows for the use of raters and rating scale response formats with more than two categories. The following simulations assumes that 3 three-category rating scales are used and that half of the respondents are being rated by two raters and the others by a single one. The only side condition (to be discussed later) is that the ratings are connected in a graph theoretical sense. Space limitations prevent deriving the equations for the general case, analogous to Equations 2 through 5. Instead, the algorithms below are supported by computer simulations rather than formal proof or derivations. Those interested in a more complete presentation should consult [8].

- **Rating scales** are modelled analogous to the binary case. Figure 3 shows that frequencies are collected in a matrix *F* with sides *nitems.nsteps1*, where *nitems* is the number of items, *nsteps* denotes the number of answer categories and *nsteps1 = nsteps –1*. Higher ratings (i.e., *1, ..*) occur along the rows, whereas lower values occur along the columns *(..., nsteps1*). Each new rating has to be compared to all others, requiring a total of *nitems(nitems – 1) / 2* comparisons. Realistic applications rarely involve over five rating scales which takes under 10 numerical comparisons. As an example, assume that the *very first* observations to be added to the zero matrix *F* are 0, 2, and 1 for items 1, 2, and 3, respectively. In this case the item table is updated as is shown by the 1's in the left table. Of course, later observations will be added cumulatively as grading progresses.
- Computing the **rater severity** parameters requires updating a second *F* matrix of size *nrat x nrat* (see Figure 3, right side) where *nrat* denotes the total number of raters – and the focus is on the sum of the ratings. In particular, whenever the same person is being rated by a second (or 3rd, 4th, ...) rater, the sum of the last ratings is compared to the earlier summed ratings. The table then simply tallies by raters the number of times one of the sums exceeds the other *by exactly one point*.
- Item biases can be studied by computing group-specific *rating x categories F* matrices (see below).

Computational Effort During Data Collection. Because the number of test takers can be very large, updating of the rater table is far more computationally expensive overall than is the updating the item table. That is, one has to keep track of doubly rated students, and their rating sums of have to be compared across raters. In the worst case, each new case may require inspecting all previously processed test-takers. Assuming that *npers* students have already been graded, this could potentially require *npers.(npers-1)/2* comparisons. The use of random access techniques can reduce this number.

Given the frequency matrices, the following steps are required to obtain actual parameter estimates:

- The ratio of off diagonal elements is required. For instance, the rater matrix *F* in Figure 3 yields the ratios $R_{ij} = F_{ji} / F_{ji}$ shown in the centre table of Figure 4. *See below for dealing with zeros*.
- Taking the logarithms of each entry in *R* yields the matrix *log(R)* shown on the right of Figure 4.
- The row-means of *log(R)* represent the severity parameters S_k (rightmost S_k column of Figure 4).
- Updating the *D* and *S* frequency matrices is computationally independent. Thus, these updates can be performed *in parallel using*

different processors thereby decreasing computation time.

- The parameters' *SE* can be obtained via bootstrapping (at least of 25 samples are required [8])
- Updating the person and rater frequency matrices (see Figure 3) shares no computational steps. To save execution time, these updates can thus be done simultaneously using parallel processing.
- The computation of the items' difficulty parameters proceeds analogously, and yield $D_{i1} = D_i + F_1$, $D_{i2} = D_i + F_2$ as row means (not shown). The convention that S $D_i = 0$ identifies the D_i and the F_w.

Figure 4: F^2 (left), R (middle) and log(R) (right) matrices for rater data in right-side of Figure 3.

	Rater 1	Rater 2	Rater 3	Rater 4
Rater 1	170	113	39	36
Rater 2	61	76	15	27
Rater 3	25	223	84	111
Rater 4	42	104	82	84

Rater 1	Rater 2	Rater 3	Rater 4
1.00	1.85	1.56	0.86
0.54	1.00	0.07	0.26
0.64	14.87	1.00	1.35
1.17	3.85	0.74	1.00

Rater 1	Rater 2	Rater 3	Rater 4	S_k
0.00	0.62	0.44	-0.15	0.23
-0.62	0.00	-2.70	-1.35	-1.17
-0.44	2.70	0.00	0.30	0.64
0.15	1.35	-0.30	0.00	0.30

Zero Entries. Above we assumed that the frequency matrices in Figures 3 contain no zeros, but this is not actually the case. However, raising *F* to the next power (i.e., $F^{p+1} = F^p.F$) will correct this by connecting indirectly connected pairs of items or raters thereby decreasing the number of zeros [8, 9]. It is almost always necessary to raise *F* to a second power, but higher powers are rarely needed. For instance, the hypothetical rater frequency matrix in the right-hand side of Figure 3 contains structural zeros along the diagonal, as well an additional zero in row 3 and col 3. As is illustrated by the left table in Figure 4, raising this matrix to the second power causes all zeros to disappear.

Computational Effort For Parameter Updates. The computational efforts needed to obtain actual parameter estimates is small. In particular, creating the *R* and *log(R)* matrices for items and raters requires computational effort proportional to $(nitems \times nsteps)^2$ and $nraters^2$, respectively. In practice, 30 > *nraters* > *nitems x nsteps*, and processing matrices of this size requires negligible effort.

PROX, JMLE [3] or a Conditional Maximum Likelihood approaches [1] can all be used to estimate students' T_j. This can efficiently be implemented as a "raw-sum to logit" lookup table for a neutral rater ($S_k = 0$), adjusting the table entry by the raters' estimated severity and averaging across raters.

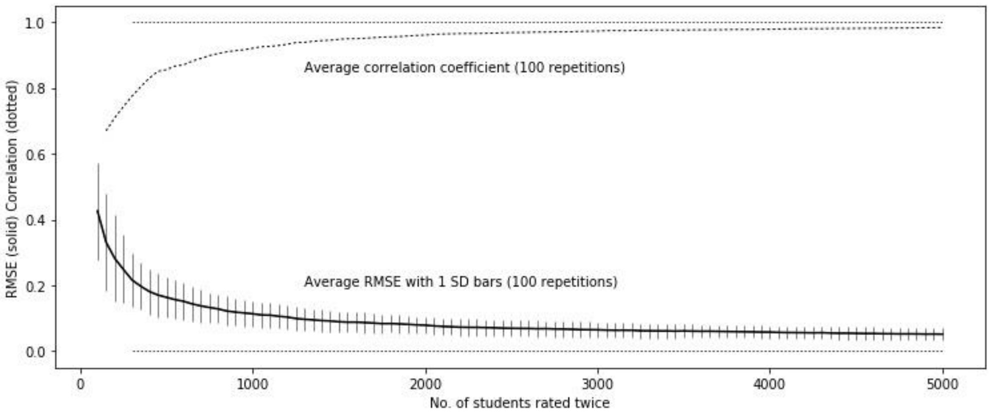
3. Simulation Results

Jupyter iPython 3.6.5 was used to simulate the grading of student essays written for tests of reading comprehension. Hypothetical essays were graded for "*completeness*" and "*style*" and these items had simulated difficulties $D_i = -0.5$ and 0.5 respectively (see Equation 1). As in Figure 1, each item used a three-category rating-scale (e.g., "*insufficient*," "*intermediate*," and "*sufficient*") with step values $F_w = \{-1.1, 1.1\}$. There are a total of five graders available to grade the essay answers of 10,000 test takers.

The graders' severity parameters S_k were drawn from *N(0,0.25),* which agrees with values obtained in actual large-scale grading applications [6]. The reading abilities of 10,000 students (T_j), were drawn from *N(0.25, 1.1)*. Half of the students were graded twice, the others once. Assignment of graders to students was random, except raters occurred equally often. While all "students" contribute to the item and steps calibrations, the rater severity estimates depend *solely* on the students that are graded twice.

Parameter Recovery. As was the case for binary items, the difficulty parameters and step-values could be recovered admirably, with the Root Mean Square Error (*RMSE*) equal to *0.04* – even after just 100 students had been graded. Items' step values likewise were recovered accurately (*RMSE = 0.02*). Although this work has not yet been completed, the accurate recovery of *D* and *F* implies that the *JMLE* recovery of the person parameters *T* promises to be accurate as well. For efficiency, estimating *T* should be combined with the computation of diagnostic and quality-control information.

Figure 5: Simulated recovery of grader severity parameters across number of students graded twice.



The main interest here is the recovery of graders' severity parameters as these require the greatest computational effort. Figure 5 shows the average *RMSE* (*Y-axis*) of the S_k estimates as a function of the number *N* of twice graded students (along *X axis*). The vertical bars reflect the standard deviation (*SD*) of these estimates over 100 independent repetitions with different samples of 10,000 "students."

It can be seen that S_k estimates' *RMSE* decreases (correlation increases) rapidly. This trend slows after *N = 300,* eventually tending to zero (one). Very similar patterns obtain for other realistic parameter settings. The present approach was highly efficient, e.g., given the updated frequency matrices, the computation of the item, steps, and grader parameters consistently took less than 0.001 seconds. The updating of the frequency matrices was quite efficient as a complete run with 10,000 students and 250 parameter updates required less than 0.5 seconds on a standard 2014 MacBook Pro laptop computer.

4. Discussion and Conclusion

We found that the present approach can estimate *MFRS* parameters without noticeable delay for plausible numbers of students and raters. For instance, current *OBJECTIVE* runs with up to 2 million students still require just 0.001 seconds to update all item, steps, and rater parameters. An alternative approach was used to recover the model parameters as the principal eigenvector of *log(R)* matrices [8, 9]. However, this did *not* improve performance or precision and the results are not further reported.

This paper did not address the estimation of the person parameters, nor did it discuss the computation of the diagnostics and quality control statistics [1, 2, 3]. Also, parallel processing as referred to above has not yet been implemented (the use of Python's *"multiprocessing"* libraries is considered for this). This work is progressing and I intend to report further results on future occasions.

References:

1. Van der Linden, W.J. & Hambleton, R.K. (1997). Handbook of modern item response theory. New York: Springer.
2. Lange, R. & Lange, X. (2012). Quality control in crowdsourcing: An Objective Measurement Approach to identifying and correcting rater effects in the social evaluation of products and services. AAAI Spring Symposium Series, North America, Stanford, CA, 2012.
3. Linacre, J. M. (1989). Many-facet Rasch measurement. Chicago: MESA Press.
4. Linacre, J. M. (2018). Facets® Rasch measurement computer program. Beaverton, OR: Winsteps.com.
5. Andrich, D., Sheridan, B. E., & Luo, G. (2004). RUMM2020: Rasch unidimensional measurement models [Computer software]. Perth, Western Australia: RUMM Laboratory.
6. Arce-Ferrer, A. & Lange, R. (2012). Assessing incidence and consequence of rater effects on open-ended scoring. Paper presented at the 8-th Conference of the International Test Commission (ITC). Amsterdam, The Netherlands, July 3-5, 2012.
7. Rasch, G. (1960/1980). Probabilistic models for some intelligence and attainment tests. (1960, Copenhagen, Danish Institute for Educational Research), 1980, Chicago: U. of Chicago Press.
8. Garner, M. & Engelhard, G.E. (2009). Using paired comparison matrices to estimate parameters of the partial credit Rasch measurement model.

Journal of Applied Measurement, 10, 30-41.
9. Saaty, T. L. (1996). The analytic hierarchy process. Pittsburgh: RWS Publications.

Sampling design proposal for Household Survey in Cambodia with provincial domain

Thy Veun[1], Erniel B. Barrios[2], Kevin Carl P. Santos[2]
[1]Deputy Director, National Institute of Statistics, Ministry of Planning, Phnom Penh, Cambodia
[2]University of the Philippines-Diliman

Abstract

A sampling design for household surveys with provinces as domain is deemed necessary because the estimates of households/person characteristics at granulated levels rather than the national and regional levels are increasingly becoming important. In this paper, a new sampling design for the household survey in Cambodia with provincial domain is proposed. In this sampling design, a further stratification within each province using the number of households in each village is introduced in addition to the sampling designs that have been used in the current practices of the household surveys in Cambodia. In this stratification, each provincial frame, which contains the list of villages, of Cambodia is stratified into five strata, and three methods are used to determine the strata boundaries for each province. These methods are (i) the generalized Lavallée and Hidiroglou (1988) method with Kozak algorithm, (ii) the cumulated root frequency method of Dalenius and Hodges (1959), and (iii) the geometric stratification of Gunning and Horgan (2004). Three levels of required precision (1%, 3%, and 5% CV) are used in conjunction with Neyman allocation to determine the sample size in each stratum and its associated boundaries. Only four among the nine sets of samples from the said stratification with the lowest sample size are finally proposed. The results show that the required sample size in each province and in the whole country is greatly reduced when compared with the case when no further stratification within each province is done. Similarly, the results from the assessment on the performance of the new proposed design using the sample data from the 2014 Cambodia Socio-Economics Survey are also found to be acceptable within a precision level.

Keywords

Stratification; generalized Lavallée and Hidiroglou method with Kozak algorithm; cumulated square root method; geometric stratification; stratum boundary

1. Introduction

A sampling design for household surveys with provinces as domain is deemed necessary because the estimates of households/person characteristics at granulated levels rather than the national and regional levels are increasingly becoming important. Cambodia Socio-Economic Survey (CSES), which has been conducted regularly every year from 2007 to 2017, and every two-year afterwards, provides such estimates only at national and regional levels. This limitation posts a lot of challenges to policy makers and other data users in their policy studies and preparations, and their related statistics compilations at their desired granulated levels. Because of the unavailability of these data, the current estimates of such statistics are based dominantly on the benchmarks from the national estimates which somehow distort the actual representation of individual provinces. Aside from these statistics, there are many other economic and social indicators for the 2030 Sustainable Development Goals which are needed to measure and compile at provincial level using data from this household survey. Being able to do so, the study on the development of alternative sampling designs for the existing household surveys which can provide the reliable estimates of such indicators at a more granulated level, specifically at provincial level, is required. This study shall contribute insights on the construction of a new sampling design for household surveys in Cambodia with provincial domains. This study aims to search for sampling techniques and statistical methods applicable on the kind of population characteristics of Cambodia.

The sampling design of the existing CSES from the 2014 to 2017 were done based on the stratified three-stage sampling design. In the stratification, the frame of all villages in Cambodia was stratified into 38 strata frames (19 province groups and urban-rural). Among the 19 province groups, 14 of which correspond to individual provinces and 5 of which correspond to grouped provinces as a combination of two adjacent provinces. From each frame, implicit stratification was also carried out before the selection of sample villages by sorting the frame according to their geographical orders. The sample size of urban villages was determined for 30% of the total sample size. For each stratum the sample size was then allocated proportionally according to the number of villages of the stratum among the whole urban or rural areas of the country.

With goal of providing provincial estimates, this study aimed to compare different methods in stratifying the population based on the distribution of the target variables of CSES. Section 2 provides description of the data used and the different stratification approaches compared in this study. Section 3 discusses the results of the study, whereas Section 4 provides some conclusions based on these results.

2. Methodology

The most updated list of all villages in Cambodia from the 2008 population census with some updates on the administrative and other related changes was used in this study. This has served as the frame for all household related survey in Cambodia so far. This frame contains the auxiliary information such as the names and codes of provinces down to villages, the location codes whether the village is in the urban or rural area within the province, the total number of households and population in each village, and the number of EAs (enumeration area) in each village. Based on the frame, there are total of 14,340 villages and 2,983,350 households across the country of 24 provinces and one capital. The average number of households (considered as the size measure in this study) per village is 208 with the minimum of 10 and the maximum of 4,158 households. The result from the study on the frame shows that 24 out of 25 provinces and capital have its skewness statistic greater than 1 (1.4 to 7.8) which implies that the number of households in each village of all provinces and capital, except Kep province with the skewness statistics of -0.04, are highly and positively skewed distributed. This indicates that the size measure characterizes a skewed distribution.

The alternative sampling design for the household survey to be proposed in the study is the stratified three-stage sampling design the same as the design in the existing survey. The difference between this proposal and the existing one is that this proposal introduced a further stratification within each province instead of individual province or grouped province. In this design, each province frame will be stratified into five strata using the number of households in each village as the stratification variable. The reason behind the decision to use five strata in each province came from the experiment on the trial uses of three, four, and five strata in each province. The results of these trials showed that five strata gain the highest precision level among the three options,and can be controllable. As a result, five strata were chosen to be further stratified in each province in this study. The strata boundaries within each province were determined using three methods as follows:

1) ***The generalized Lavallée and Hidiroglou method with Kozak algorithm (LHK):*** This iterative method was used to determine the optimal strata boundaries of the stratifying variable *X*, the total number of households in each village, by accounting for the stratum's relative weight, mean, and variance. In this method, the minimum sample sizes per stratum were calculated under three precision levels of 1%, 3%, and 5% CV (coefficient of variation), and under Neyman Allocation rule. To determine the stratum boundaries, the population in each province was sorted with respect to the values of stratifying variable in the ascending order, $x_0, \ldots, x_L$ where x_0 is the minimum value and x_L is the maximum value of stratifying variable. The following steps were followed:

Step 1: Start with some arbitrary boundaries such that $x_0 < x'_1 < \cdots < x'_{L-1} < x'_L$, with the constraint $N_h \geq 2; h = 1, \ldots, L$; and $2 \leq n_h \leq N_h, h =$

$1, \ldots, L-1$, where n_h was determined with Neyman Allocation as: $n_h = \frac{nW_hS_h}{\sum_{h=1}^{L-1} W_hS_h}$ (Eq. 1), are satisfied.
Step 2: Calculate $W'_h, \bar{X}'_h,$ and $S_h^{2\prime}$ for these boundaries, for $h = 1, \ldots, L-1$.
Step 3: Replace the initial sets of boundaries by $x_1'', \ldots, x_{L-1}''$,
where $x_h'' = \frac{-\alpha_h' + \sqrt{\beta_h'^2 - 4\alpha_h'\gamma_h'}}{2\alpha_h'}$, $h = 1, \ldots, L-1$(Eq. 2) (see more details in Lavallée and Hidiroglou (1988, p.38)).
Step 4: Repeat steps 2 and 3 until two consecutive sets are either identical or differ by negligible quantities.
Kozak algorithm was also used in conjunction with this method in the evaluation of each stratum's sample size as in Eq. (1), and to control for the constraints as stated in step 1. Kozak algorithm was used in this case because it often provides smaller sample size requirement than others, such as Sethi algorithm, in the case of this studied population.
2) The Cumulated Root Frequency Method of Dalenius and Hodges (CDH): This non-iterative was also used to calculate the optimum stratum breaks in the values of the stratification variable. Because there is no precise rule concerning the optimal number of classes, the Sturges' formula was used as the approximation to the number of classes *K*. Similar to LHK, the determination of the minimum sample sizes within each stratum were calculated basing on the same specified levels of precision and allocation rule together with the determination of stratum bounds.
3) The Geometric Stratification Method of Gunning and Horgan (GGH): This method is one of the noniterative methods that is extremely convenient in the implementation of finding the stratum breaks using only the minimum and maximum values of stratification variable to obtain the boundaries. According Gunning and Horgan (2004), the stratum bounds are the terms of geometric progression: $x_h = ar^h$, for $h = 0,1,2, \ldots, L$, where $a = x_0$ *a* (the minimum value of number of households within the village of respective province), and $ar^L = x_L$ (the maximum value of number of households within the village of respective province).

The stratification package in R was used as an aid in finding the optimum stratum boundaries and the minimum sample sizes in each stratum as well as the overall sample size of each province under the same specified levels of precision, and under the same allocation rule. The take-none and take-all strata were not set for LHK.

Because the sampling design was done through a multistage sampling design rather than a direct element sampling, the effect between this complex sampling and a simple random sampling schemes shall be taken into account in the sample size determination. The design effect (*deff*) was used to account for this effect. For this accounting, the preliminary sample sizes calculated from the three methods above were multiplied by the design effect of 2 in this study. The *deff* of 2 was used because the study on *deff* in each province on some indicators from the CSES2014, such as, the number of household in each village, total household income and consumption, showed that the number of households in each village exhibited *deff* ranging from 0.3 to 1.8 while total expenditure and total income variables have their *deff* ranging from 0.5 to 1.7 and from 0.5 to 1.8 respectively.

3. Main finding/results

The required sample sizes of households in each province with a further stratification within each province and the without-further stratification are shown in the Table 1. Results show that within 5 percent level of precision, very huge reduction in the required sample sizes of households for each province was achieved in the case when a further stratification within each province was done compared to the one without a further stratification. Without a further stratification, the total sample size of 42,515 households is required for whole country, whereas only 8,340, 10,290, and 14,670 households are required, respectively, for LHK, CDH, and GGH under the case when a further stratification within each province was done. Among these three methods, LHK method provided the smallest sample size of households followed by the CDH method. The GGH method of stratum boundary determination gave the largest sample size. To gain more insight into how these three methods have worked for this studied population, all the three levels of required precision used are showed in the Table 2. It was found that LHK was still the method that provided the smallest sample size among the three, followed by the CDH method, in all levels of required precision. The GGH method performed the least because it provided the largest sample size requirement in all required levels of precision.

Table 1. The required sample sizes of households in each province under 5% CV without and with further stratification within province

Province Name	Total number of households	Without further stratification within province	With further stratification within province		
			LHK	CDH	GGH
Banteay Meanchey	153,396	2,032	420	450	750
Battambang	226,460	1,233	330	330	630
Kampong Cham	222,380	1,094	300	300	570
Kampong Chhnang	107,875	1,099	300	390	540
Kampong Speu	157,207	851	270	330	480

Province Name	Total number of households	Without further stratification within province	With further stratification within province		
			LHK	CDH	GGH
Kampong Thom	144,539	1,055	300	360	570
Kampot	134,871	768	210	270	450
Kandal	237,056	1,515	390	450	810
Koh Kong	25,308	3,428	420	450	570
Kratie	68,234	1,192	330	360	570
Mondul Kiri	13,150	1,745	300	390	450
Phnom Penh	267,500	2,817	660	960	1,170
Preah Vihear	44,368	3,270	420	660	660
Prey Veng	243,701	1,117	300	390	630
Pursat	91,099	1,106	270	300	420
Rattak Kiri	32,095	2,456	480	540	840
Siem Reap	186,393	2,522	420	540	840
Preah Sihaknouk	43,179	3,217	390	450	480
Stung Treng	23,703	1,923	270	390	330
Svay Rieng	124,193	856	210	270	390
Takeo	194,113	955	270	330	570
Otdar Meanchey	46,341	1,357	300	390	540
Kep	8,568	516	150	150	150
Pailin	15,650	3,039	300	390	480
Thbong Khmum	171,980	1,352	330	450	780
Total	**2,983,350**	**42,515**	**8,340**	**10,290**	**14,670**

Among the three levels of required precision, the level of 1% CV required too large sample size in all methods which would be unaffordable to conduct due to budget constraints. The existing survey has ever been conducting with up to 12,096 households only in the biggest sample year. Having considered both the acceptable levels of precision and the cost of conducting it, only four among the nine sets of samples from the said stratification with the lowest sample sizes are finally proposed for the sampling design of household survey in Cambodia with provincial domains. These four sets of samples were the samples from the LHK and CDH methods with the 3 and 5% levels of required precision. These four sets of sampled villages were selected by the systematic probability proportional to size (sys-PPS) sampling scheme using the number of households in each village as the size measure. These selected samples were assessed by using the sample data on the employment rate and adult literacy rate from the 2014 CSES. Table 3 illustrated about the results of assessments on the performance of the two sets of samples from the LHK and CDH methods with 5% CV. The assessment indicated that the two sets of samples performed well within the acceptable levels of required precision. LHK5% produced the estimate with 100% probabilities that the estimated CVs will fall within less than 5% in 23 out of 25 provinces and capital for employment rate, and in 20 out of 25 provinces and capital for adult literacy rate. Similarly, CDH5% produced the estimate with 100% probabilities that the estimated CVs will fall within less than 5% in 23 out of 25 provinces and capital for both employment rate and adult literacy rate. The performance assessment of the samples by LHK and CDH methods with 3% CV shown that all the 25, and 24 out of 25 provinces and capital have their estimates with 100% probabilities of estimated CV falling within less than 5% respectively of employment rate adult literacy rate.

Table 2. The required sample sizes of households in each province with three levels of precision under LHK, CDH, and GGH methods

Province Name	LHK			CDH			GGH		
	1% CV	3% CV	5% CV	1% CV	3% CV	5% CV	1% CV	3% CV	5% CV
Banteay Meanchey	4,410	1,020	420	5,100	1,110	450	7,650	1,770	750
Battambang	4,830	780	330	5,310	840	330	8,010	1,530	630
Kampong Cham	4,680	720	300	5,070	810	300	7,770	1,350	570
Kampong Chhnang	3,600	750	300	3,870	930	390	6,210	1,290	540
Kampong Speu	4,650	660	270	4,890	780	330	7,530	1,140	480
Kampong Thom	4,530	720	300	4,800	840	360	7,650	1,440	570
Kampot	3,120	510	210	3,420	630	270	5,160	1,020	450
Kandal	6,000	960	390	6,060	1,020	450	10,050	1,980	810
Koh Kong	1,410	720	420	1,980	810	450	2,160	900	570
Kratie	2,550	690	330	2,970	750	360	4,140	1,260	570
Mondul Kiri	1,230	510	300	1,530	600	390	1,860	750	450
Phnom Penh	5,970	1,590	660	8,550	2,040	960	9,750	2,640	1,170
Preah Vihear	2,280	750	420	2,610	930	660	4,200	1,410	660
Prey Veng	4,920	750	300	5,310	900	390	9,120	1,530	630
Pursat	3,510	660	270	3,810	750	300	5,190	1,020	420
Rattak Kiri	2,460	870	480	4,140	1,170	540	4,500	1,710	840
Siem Reap	4,410	1,020	420	6,510	1,350	540	9,360	2,010	840
Preah Sihaknouk	1,320	630	390	2,130	720	450	1,950	780	480
Stung Treng	750	600	270	2,040	660	390	2,280	690	330
Svay Rieng	3,690	570	210	3,960	630	270	5,190	870	390
Takeo	4,710	690	270	4,830	780	330	8,610	1,440	570
Otdar Meanchey	2,730	690	300	3,000	900	390	4,230	1,170	540
Kep	360	150	150	390	210	150	420	240	150
Pailin	1,140	570	300	1,800	660	390	1,710	870	480
Thbong Khmum	4,950	870	330	5,520	1,080	450	9,300	1,920	780
Total	84,210	18,450	8,340	99,600	21,900	10,290	144,000	32,730	14,670

4. Conclusion

The results suggest that further stratification within the province (in this case, five strata were stratified within the province) greatly reduce the sample size requirements for those provinces as well as the whole country. Among the three methods of the stratum boundary determination, LHK provided the boundaries with the smallest sample size requirement, whereas the boundaries of GGH requires the largest sample size. Only the sample sizes at the required levels of three and five-percent CVs are considered feasible in practice since it is too costly to consider the sample sizes produced by one-percent CVs. In the estimation of rates for the common and the less common characteristics such as employment rate and adult literacy rate with the proportion as low as 40%, both the LHK and CDH methods at the required level of five-percent CV are acceptable in terms of both cost and precision. One can choose the CDH method if precision is more important than cost. The sets of sample under the CDH method generally exhibit higher precision in the estimation than those under the LHK method, but with larger sample size than those of the LHK method. If the survey variables are highly correlated with the number of households in each village, the sampling designs that have been proposed in this study can be considered. For multipurpose household survey, such as the CSES, multivariate stratification shall be incorporated through the application of multi-stratification variables in the further stratification within the provinces if such variables are available for all elements in the studied population. Use different variables for the stratification variable and the size measure if such variables are available for all elements in the studied population to ensure the appropriate gain from these survey techniques.

Table 3. Distributions of CVs of Estimated Employment Rate and Adult Literacy Rate by Provinces

Province Name	Distribution of Estimated Coefficient of Variation (CV)							
	LHK 5% Samples				CDH 5% Samples			
	Employment Rate		Literacy Rate		Employment Rate		Literacy Rate	
	< 5%	5-10%	< 5%	5-10%	< 5%	5-10%	< 5%	5-10%
Banteay Meanchey	100	-	100	-	100	-	100	-
Battambang	100	-	100	-	100	-	100	-
Kampong Cham	100	-	100	-	100	-	100	-
Kampong Chhnang	100	-	100	-	100	-	100	-
Kampong Speu	100	-	96.9	3.1	100	-	100	-
Kampong Thom	100	-	100	-	100	-	100	-
Kampot	87.5	12.5	87.5	12.5	92.2	7.8	98.4	1.6

Province Name	Distribution of Estimated Coefficient of Variation (CV)							
	LHK 5% Samples				CDH 5% Samples			
	Employment Rate		Literacy Rate		Employment Rate		Literacy Rate	
	< 5%	5-10%	< 5%	5-10%	< 5%	5-10%	< 5%	5-10%
Kandal	100	-	100	-	100	-	100	-
Koh Kong	62.5	37.5	100	-	75	25	100	-
Kratie	100	-	100	-	100	-	100	-
Mondul Kiri	100	-	100	-	100	-	100	-
Phnom Penh	100	-	100	-	100	-	100	-
Preah Vihear	100	-	65.6	34.4	100	-	79.7	20.3
Prey Veng	100	-	100	-	100	-	100	-
Pursat	100	-	100	-	100	-	100	-
Rattak Kiri	100	-	65.3	34.7	100	-	100	-
Siem Reap	100	-	100	-	100	-	100	-
Preah Sihaknouk	100	-	100	-	100	-	100	-
Stung Treng	100	-	100	-	100	-	100	-
Svay Rieng	100	-	100	-	100	-	100	-
Takeo	100	-	100	-	100	-	100	-
Otdar Meanchey	100	-	100	-	100	-	100	-
Kep	100	-	100	-	100	-	100	-
Pailin	100	-	100	-	100	-	100	-
Thbong Khmum	100	-	87.5	12.5	100	-	100	-

References

1. Baillargeon, S., and L.-P. Rivest (2015). Univariate Stratification of Survey Populations. *R Package*, available on the CRAN website at http://www.r-project.org/. 2015.
2. Cochran, W.G. (1977). Sampling Techniques. Third Edition. New York: John Wiley & Sons, Inc.
3. Dalenius, T., and Hodges, J.L., Jr. (1959). Minimum variance stratification. Journal of the American Statistical Association, 54:88-101.
4. Gunning, P., and Horgan, J.M. (2004). A new algorithm for the construction of stratum boundaries in skewed populations. Survey Methodology, 30:159-166.
5. Kozak, M. (2004). Optimal stratification using random search method in agricultural surveys. Statistics in Transition, 6:797-806.
6. Lavallée, P., and Hidiroglou, M. (1988). On the stratification of skewed populations. Survey Methodology (Statistics Canada), 14:33-43.
7. Lohr, S. (2010). Sampling: Design and Analysis. Second Edition, Brooks/Cole, Cengage Learning, USA.

Model-assisted estimation using panel data and data from independent samples for the Philippine Family Income and Expenditure survey

Erniel B. Barrios[1], Divina Gracia L. del Prado[2], Mae Abigail C. Oberos[2], Desiree Robles[2], Jo Louise Buhay[2]
[1]School of Statistics, University of the Philippines
[2]Statistical Methodology Unit, Philippine Statistics Authority

Abstract
The domains for the Philippine Family Income and Expenditure Survey (FIES) are the regions. Design-unbiased estimation for the provinces (smaller areas) would necessitate an increase the sample size to about four times. Considering further that FIES is conducted semi-annually (at the end of reference semester), using the same sample households for the survey year and administered with a lengthy questionnaire, response burden and refusals increases along with problems with willingness of interviewers to complete the task for the entire survey duration. This problem is addressed by keeping the regions as survey domains but partially rotating the samples in the two rounds of data collection. This study then proposes a model-assisted estimation procedure in combining panel data and data from independent samples to generate annual estimates at the provincial level. Simulation studies indicate the advantages of model-assisted estimation procedure over design-unbiased method in generating estimates at the provincial level. Furthermore, sample size requirement using model-assisted approach is only about half of the required sample size for the design-unbiased method for reliable provincial level estimates.

Keywords
Rotated sample; skewed population; cut-off sampling

1. Introduction

Nationwide surveys intended to generate official statistics often incur huge cost and complex logistical issues in survey operations, hence, often constraint statistical organizations to consider bigger areas (in some cases, national) as the domains. However, more granular data is also needed to monitor progress among countries towards achievement of sustainable development goals. The Philippine Family Income and Expenditure Survey (FIES) is the source of many indicators for various goals, the most important of which are poverty-related indicators. FIES has always been a panel survey where same households are visited every semester. While this has been observed to generate "reliable" data for income and expenditures, it easily causes respondents' fatigue specially since it also uses a lengthy questionnaire with very complicated response structure. In the last survey round (2015), sample size that can generate reliable estimates up to regional level is 45,000 households. With similar design, to generate provincial level estimates would require 180,000 households to be visited every semester.

This paper uses the 2015 FIES data in simulating a partially rotated sample to evaluate a proposed model-assisted estimation procedure that combines panel data and data from independent samples to generate estimates at provincial level with desirable efficiency level.

2. Methodology

Implementing panel and independent samples combination

This paper will make use of rotated samples (partly independent, partly panel) in a repetitive survey (2 rounds). This method uses a combination of panel data and independent samples, i.e., part of the original samples from the first round are visited again in the second round of the survey. Rotating samples has been used as a sampling strategy when estimates are produced regularly over time. In this method, equal-sized sets of sample units are brought in and out of the sample in a specified pattern. Partially rotated samples are intended to reduce variance of estimators, reduce the survey costs associated with completely independent samples, and minimize respondent's fatigue and refusals in a panel survey.

(Graham, 1963), considered the implied population in a partially rotated sample to be split into unmatched (that portion where the independent samples were drawn) and matched (where the panel sample were drawn). 2015 FIES data is split into the matched and unmatched dataset denoted as $\hat{\hat{Y}}_{1,u}$ and $\hat{\hat{Y}}_{1,m}$ for 1st vist and $\hat{\hat{Y}}_{2,u}$ and $\hat{\hat{Y}}_{2,m}$ for 2nd visit. Samples are matched at the primary sampling unit (PSU) and secondary sampling unit (SSU) level. The PSUs are enumeration areas (EA) and SSU are the households (HH). In a domain, half of the sample PSUs are matched for Rounds 1 and 2. In matched PSU, same HH are visited in Rounds 1 and 2. Two sets of estimates (estimates with corresponding variance) will be generated from each visit, given in the following:

	$\hat{\hat{Y}}_1$	$\hat{\hat{Y}}_2$
Unmatched samples	$\hat{\hat{Y}}_{1u}$ $V\left(\hat{\hat{Y}}_{1u}\right) = \frac{E\sum_{i=1}^{E} H_i - u\sum_{j=1}^{u} u_j}{\left(E\sum_{i=1}^{E} H_i\right)\left(u\sum_{j=1}^{u} u_j\right)} \frac{S_{1u}^2}{u\sum_{j=1}^{u} u_j}$	$\hat{\hat{Y}}_{2u}$ $V\left(\hat{\hat{Y}}_{2u}\right) = \frac{E\sum_{i=1}^{E} H_i - u\sum_{j=1}^{u} u_j}{\left(E\sum_{i=1}^{E} H_i\right)\left(u\sum_{j=1}^{u} u_j\right)} \frac{S_{2u}^2}{u\sum_{j=1}^{u} u_j}$
Matched Samples	$\hat{\hat{Y}}_{1m}$ $V\left(\hat{\hat{Y}}_{1m}\right) = \frac{E\sum_{i=1}^{E} H_i - m\sum_{j=1}^{m} m_j}{\left(E\sum_{i=1}^{E} H_i\right)\left(m\sum_{j=1}^{m} m_j\right)} \frac{S_{1m}^2}{m\sum_{j=1}^{m} m_j}$	$\hat{\hat{Y}}_{2m}$ $V\left(\hat{\hat{Y}}_{2m}\right) = \frac{E\sum_{i=1}^{E} H_i - m\sum_{j=1}^{m} m_j}{\left(E\sum_{i=1}^{E} H_i\right)\left(m\sum_{j=1}^{m} m_j\right)} \frac{S_{2m}^2}{m\sum_{j=1}^{m} m_j}$

where E= total number of PSUs; e – number of sample PSUs; H_i- total number of HH in i^{th} PSU; h_i – total number of sample households in i^{th} sample PSU, and; y_{ij} – measurement of indicator for j^{th} sample household in i^{th} sample PSU.

Modelling

We used the estimates obtained from the matched and unmatched samples to generate estimates of annual totals for the indicators. The annual estimate of the mean and the corresponding variance is

$\hat{\hat{Y}}_2^* = \delta_1 \hat{\hat{Y}}_u + (1-\delta_1)\hat{\hat{Y}}_{2m}^*$ with variance $V\left(\hat{\hat{Y}}_2^*\right) = \delta_1^2 V\left(\hat{\hat{Y}}_u\right) + (1-\delta_1)^2 V\left(\hat{\hat{Y}}_{2m}^*\right)$ where components are computed using the ff: $\gamma_1 = \frac{V(\hat{\hat{Y}}_{2u})}{V(\hat{\hat{Y}}_{1u})+V(\hat{\hat{Y}}_{2u})}$; $\hat{\hat{Y}}_u = \gamma_1 \hat{\hat{Y}}_{1u} + (1-\gamma_1)\hat{\hat{Y}}_{2u}$;$V\left(\hat{\hat{Y}}_u\right) = \gamma_1^2 V\left(\hat{\hat{Y}}_{1u}\right) + (1-\gamma_1)^2 V\left(\hat{\hat{Y}}_{2u}\right)$; $\hat{\hat{Y}}_{2m}^* = \hat{\hat{Y}}_{2m} + b\left(\hat{\hat{Y}}_1 - \hat{\hat{Y}}_{1m}\right)$; $V\left(\hat{\hat{Y}}_{2m}^*\right) = \frac{S_{2m}^2(1-b^2)}{m}$ and $\delta_1 = \frac{V(\hat{\hat{Y}}_{2m}^*)}{V(\hat{\hat{Y}}_u)+V(\hat{\hat{Y}}_{2m}^*)}$.

Assessment

The reliability of estimates per domain and per disaggregation is evaluated using the Coefficient of Variation (CV). Efficiency of an estimator is considered acceptable if CV is lower than 10%.

Simulation exercise

We simulate various rotation patterns and assess its effect on CV of provincial level estimates of different indicators. We vary the composition share of matched and unmatched samples for the following cases: 40% matched and 60% unmatched, 75% matched and 25% unmatched and 90% matched and 10% unmatched.

3. Results

Provincial Estimates of Total/Mean Income and Total/Mean Expenditure by Income Class, Expenditure Class, Family Size

The use of the matched and unmatched samples generally yield reliable estimates reliable for smaller areas (85 provinces and highly urbanized cities) while retaining the sample size of 45,000 per semester visit of the respondents. At the national level, for overall mean or total income and expenditure estimates, all provinces exhibited estimates with CV not exceeding 10%. Disaggregation income class and expenditure class, estimates in all provinces in each class also posted reliable CVs. When disaggregated by family size, some provinces posted CVs more than 10% CV but not more than 20% CV. For one-person family group, 23 provinces had estimates with CV higher than 10%. For two-person family, only 11 provinces yield estimates with CV higher than 10%. Among three-person and four-person family, only 10% of provinces have CV exceeding 10%.

The FIES is the main source of details of expenditure used in poverty analysis and updating of consumer price index. Disaggregation to major expenditure group and its sub-items are deemed important. We computed for the total and/or mean expenditure over major expenditure groups and the proposed methodology consistently provides reliable estimates up to provincial level.

Expenditure Group	No. of Provinces/HUC with CV<10%
Food	84*
Alcoholic Beverages and Tobacco	80*
Clothing and Footwear	79*
Housing, Water, Electricity, Gas and Other Fuels	84*
Furnishing and Routine Household Maintenance	79*
Health	26*
Transport	76*

Expenditure Group	No. of Provinces/HUC with CV<10%
Communication	76*
Recreation and Culture	51*
Education	52*
Miscellaneous Goods and Services	79*
Durable Furniture and Equipment	-
Special Family Occasions	39*
Other Expenditure	26*
Other Disbursements	16*

For the food subgroups, the proposed methodology can still generate provincial estimates with reliable CVs.

Expenditure Group	No. of Provinces/HUC with CV<10%
Food	84*
Bread	84*
Meat	84*
Fish and Seafood	84*
Milk Cheese and Eggs	81
Oils and Fats	82

Expenditure Group	No. of Provinces/HUC with CV<10%
Fruit	84
Vegetables	84
Sugar, Jam, Honey, Chocolate and Confectionery	83
Food Products Not Elsewhere Classified	83
Coffee Tea and Cocoa	84
Mineral Water, Softdrinks, Fruit and Vegetable Juices	76

Provincial Estimates over Income Deciles (Implicit Stratification)

Since matched and unmatched samples from each visit are assigned independently, estimates over income deciles will be altered. Our solution if to implicitly stratification sample selection over income classes. The estimates over per capita income should be done after obtaining the annual estimate of the income per household to ensure that the per capita income decile to which sampled household belong to its true income decile.

Provincial Estimates for Median Income (Bootstrap Sampling)

In computing for the annualized model-assisted estimates, both the estimate and the variance are needed. Using the conventional method of obtaining the median estimate, generation of annualized model-assisted estimate for the median is not possible since generation of the variance of median is computationally intensive. As a solution, bootstrap resampling was done. Bootstrap resampling allows the computation of the standard error and the coefficient of variation for the estimates of the median. It also minimizes the bias and improves the reliability of estimates (Mooney and Duval, 1993) for the estimates over per capita income decile. Aside from computing the bootstrap estimates for the median, the bootstrap estimates for the annualized overall income and expenditure and design-based mean income over per capita income decile were also obtained. The purpose of obtaining the estimates for the overall income and expenditure is to verify robustness of the bootstrap estimation procedure. The expected value for the overall income and expenditure is that its annualized bootstrap estimates were close to the original annualized model-assisted estimates.

4. Concluding notes

Aggregated model-assisted annualized estimates for provinces in partially rotated samples coincide with design-unbiased estimates at the domains (regions). The CVs at the national and regional level for the overall income and expenditure as well as CVs at the national and regional level for income and expenditure over income class were all less than 10%. Moreover, provincial estimates by different disaggregation groups also yield reliable estimates. From simulation studies, increasing the percentage of matched samples does not necessarily affect improve efficiency of the estimates, but CV decreases with increasing matched sample size. Meanwhile for the estimate per capita income decile, it is recommended that the sample be selected with stratification per income class or income decile. On the estimation of median family income by per capita income decile, bootstrap method is recommended. It is proposed to use the rotated sample to produce reliable provincial estimates while retaining the sample size of the Family Income and Expenditure Survey.

References

1. Graham, John Elwood. 1963. "Rotation designs for sampling on successive occasions" Retrospective Theses and Dissertations. 2384. https://lib.dr.iastate.edu/rtd/2384
2. Mooney, C.Z. and Duval R.D., 1993, Bootstrapping A Nonparametric Approach to Statistical Inference, Sage Publications, Inc.

The Early-Warning System

Sophie Limpach, August Götzfried, Håkan Linden, John Verrinder
European Commission, Luxembourg

Abstract

Economic decisions of Multinational Enterprise Groups (MNEs) regarding their place of incorporation, ownership of their assets and economic control may result in sizeable revisions of European business statistics, National Accounts, and Balance of Payment statistics. Recording such globalisation events in a timely and consistent way across all the EU Member States concerned and ensuring comparable and consistent European statistics is a serious challenge. A close coordination between the Member States affected by such globalisation events and Eurostat is therefore needed. For this purpose, the Early-Warning System (EWS) was set up within the European Statistical System (ESS) as a light and easy-to-manage network. The system should lead to a consistent statistical treatment of major restructuring/globalisation cases of MNEs in European statistics. So far, more than 10 specific restructuring cases have been treated in the EWS. Based on the experiences gathered, the system will be further developed in 2019/2020 (with a better infrastructure and more methodological guidance).

Keywords

Globalisation; Multinational Enterprise Groups; Restructurings

1. Introduction

Major restructuring events of MNEs in 2016 and 2017 challenged the quality and consistency of European business statistics, National Accounts and Balance of Payment Statistics. As a consequence, Eurostat developed the first ideas for establishing an EWS, which were then discussed in several ESS groups, such as the Joint task force of directors of business and macroeconomic statistics on globalisation.

The European Statistical System Committee in its meeting of February 2018 supported the setting up of the EWS as a light developing system based on voluntary co-operation and learning by doing and fully respecting the Regulation on European statistics [Regulation (EC) 223/2009] on the exchange of confidential data. Several EU Member States underlined the importance of enhancing the cooperation in this regard with the European System of Central Banks (ESCB) especially at the national level.

The EWS very much relies on an established network of national EWS correspondents, coordinated by Eurostat (which ensures the secretariat), who started working on concrete restructuring cases. Furthermore, the cooperation with the ESCB and the European Central bank (ECB) has been established.

2. Purpose of the EWS

The EWS was established to detect in a timely manner and being prepared for restructuring events of MNEs. In particular, the purpose of the EWS is:

1. To facilitate an early exchange of information on restructuring cases across EU Member States directly concerned with the aim of achieving, to the extent possible, an agreed methodological treatment of the cases by the national data compilers.
2. To ensure the consistency of European statistics as regards such globalisation events.
3. To ensure a coordinated timing in the publication of first results and revisions.
4. To ensure a timely, consistent and interlinked communication towards users of national and European statistics.

The system should focus on the consistent statistical treatment of single major globalisation cases.

3. Events subject to the EWS

The EWS should be triggered in the following cases:

1. The restructuring of one or more MNEs effects not only one EU Member State, but also two or more. Restructuring of MNEs means that an MNE changes the group structure or distribution of its business model across countries (e.g. creation of new branches or other entities, the transfer of assets across borders, changes of global production arrangements such as contract manufacturing).
2. The restructuring has sizeable effects on national and on European statistics. This should be judged looking at the published data. Special attention needs to be given to National Accounts main aggregates, the Principal European Economic Indicators (PEEIs) and similar data with a high visibility.
3. Since the EU Member States concerned might be very different in economic size and structure, a single quantitative threshold for triggering the EWS is not appropriate. When dealing with restructuring cases, medium-sized and big countries should also consider the effects on smaller countries, in particular when they already know that such countries are affected.
4. The impact on the data published should also be considered in the light of the size of normal revisions for the statistics in question. Moreover,

breaks in unadjusted, calendar adjusted and/or seasonally adjusted time series as well as revisions covering longer time intervals than usually in the unadjusted data should also trigger the system.

5. If the restructuring effects on official statistics take place over a longer time, the total size of the effects should be taken into account – not just the effects on single data releases.
6. The granularity of the data published as European statistics should be taken into account (e.g. the breakdown of the data published according to the NACE Rev. 2 classification).
7. The focus of the EWS should be on the restructuring within MNEs especially since these are more difficult to detect and information on them is more difficult to obtain.
8. Restructuring between MNEs (e.g. mergers and acquisitions) might also be dealt with in the EWS.
9. Triggering the EWS by a national EWS correspondent is also possible for cases that come to the correspondent's attention but do not directly affect his/her own country.
10. Generally, the EWS network may be consulted whenever national data compilers are in doubt regarding the correct treatment of specific restructuring cases.

4. Information to be exchanged in the EWS

EWS correspondents should trigger the EWS as soon as they learn about the restructuring of an MNE with possible sizeable effects on national and European statistics. National data compilers should not wait until all facts about such cases are available. The EWS could also be triggered by information available at European level, for example by Eurostat or the ECB.

When triggering the EWS, national data compilers should attempt to provide the following information:

1. A detailed description of the restructuring case (the description could be drafted in an anonymised, stylized form "Company A opens a new branch in country X ...").
2. The statistical domains/indicators affected (e.g. Short-term business statistics, Structural business statistics, Balance of Payments, National Accounts, ...).
3. The industries (NACE) affected (if possible for confidentiality reasons)
4. The estimated effects on the data (by statistical domain, if known).
5. Other countries that might be affected (if known).
6. Timing of the restructuring (when it took/ will take place).
7. The level of sensitivity (e.g. is the information based on a newspaper article, has the concerned MNE agreed to the sharing of information, ...).

8. Information on other data compilers already contacted (e.g. National Central Banks).
9. A methodological assessment and proposal for statistical treatment (if already known).
10. Eventual plans for communication to users.

If some information listed above is not available, this should not delay triggering the EWS. In general, only descriptive information on the restructuring cases should be exchanged. A regular exchange of confidential data is not the purpose of the EWS.

5. Statistical confidentiality and the EWS

The application of existing statistical methods does not automatically guarantee a consistent treatment of restructuring cases by all data compilers concerned. These restructuring cases are often unique and might require different statistical treatment, Moreover, not all EU Member States affected might have the same information about a case. Yet, to achieve consistency in methodology and dissemination, often only a limited amount of descriptive information might be necessary. In particular, it will not be necessary to establish a regular exchange of quantitative confidential data of individual businesses.

The following principles will guide the exchange of the EWS information:

1. The EWS information shared is limited to what is necessary to achieve the EWS purpose.
2. The EWS information on a specific case is only shared among Eurostat and the national EWS correspondents of the countries affected by the specific case.
3. The national data compiler who triggers the EWS also indicates the sensitivity of the case (in particular whether the concerned business has agreed to share data, whether the information is publicly available etc.).
4. Information related to EWS cases are shared via a dedicated interest group in the secure CIRCABC environment or encrypted emails. The CIRCABC [see CIRCABC (EC)] is a collaborative platform that provides a secured working area to share information in the European Commission and among registered users, organised in interest groups. In the secure CIRCABC interest group for the EWS, a separate folder for each restructuring case has been implemented, with access limited to the actors involved.

6. The EWS process

The basic steps of the EWS process are as follows:

1. A national EWS correspondent informs the EWS secretariat at Eurostat of a restructuring case affecting national or European statistics.
2. The EWS secretariat, the EWS correspondent from the EU Member State triggering the EWS and the other EU Member States concerned form an ad-hoc task force to discuss the case. This Task Force will express a first opinion on the case and possible ways to treat it in the national and European statistics concerned.
3. Eurostat and the national statistical compilers concerned coordinate the dissemination of the statistical results.
4. In close cooperation with the concerned data compilers, Eurostat will draw up a methodological note (containing an anonymised summary of the case and methodological treatment).
5. This methodological note is stored in the secure EWS CIRCABC interest group and accessible for all EWS correspondents. In this way, a repository of cases is built up to guide the decisions on new cases.
6. Other EWS correspondents might express their opinion on this methodological note. They may also inform Eurostat that they are also affected by the case in question.
7. If a globalisation event has major effects on the data published, Eurostat releases a short explanatory note at the moment of the data release on its website.

The national EWS correspondents should ensure the involvement of their respective National Central Banks (NCBs) statistics department (counterparts) in EWS cases where external statistics, financial statistics or any underlying primary data under their responsibility are affected. Likewise, NCBs should inform the national EWS correspondents when they become aware of potential cases. The relevant NCBs and the ECB statistics department are participating fully in the case-specific task forces where external, financial statistics or any underlying statistics under their responsibility are affected.

7. The EWS cases

At the time of drafting this paper, 14 restructuring cases were reported to the EWS, and 18 EU Member States and 2 EFTA countries have been directly involved in at least one EWS case. 7 cases have been closed, and the other 7 are ongoing, i.e. information is being collected and/or the statistical impact is being assessed. 3 of the 7 ongoing cases are recent and concern a financial services group, an automotive manufacturer, and an electronic goods manufacturer.

Table. Overview of EWS cases

N°	MNE involved	Description
1	Online retailer	Former logistics centre became retail
2	Online media content seller	Relocation of an unit
3	International restaurant chain	Relocation of an unit
4	Low cost carrier	Set up of affiliate company
5	Social network	Changes to booking of advertisement
6	Oil and gas (1)	Merger
7	Oil and gas (2)	Acquisition
8	Consumer goods manufacturer	Simplification of management structure
9	Tobacco producer	Setting up of a new factory
10	Financial services group	Moving headquarters
11	Pharmaceutical company	Acquisition
12	Electronic goods manufacturer	Moving headquarters
13	High Technology manufacturer	Transfer of registration of licenses
14	Automotive manufacturer	Relocation of production

The individual cases are discussed with the national correspondents of the concerned EU Member States and EFTA countries, and the anonymised methodological notes drafted and made available to all the national EWS correspondents.

Based on the outcome of the discussions of the specific restructuring case with the national EWS correspondents concerned, Eurostat might also draft and publish an (anonymised) explanatory note, normally, at the moment of the data release. For an example of such an explanatory note, see [Eurostat 2017].

8. Additional actions and data sources for detecting MNE restructuring cases

So far, the number of cases dealt with in the EWS is quite limited. In order to get more restructuring cases treated by the EWS, the following additional actions are being studied.

- Working more closely together with national Large Cases Units (LCUs). Many EU Member States have established (or will establish) national LCUs. A close cooperation of the EWS correspondents with these LCUs is encouraged. If needed, national alert processes should be established from the LCUs to the EWS national correspondents for this purpose.

- Using web scraping for detecting MNE restructuring cases. The internet provides various options and sources of information to search for mergers and acquisitions of companies. Targeted web search of companies, its competitors, industry related information in which the company operates or demographic and economic factors could be made.
- Exploiting websites and databases dedicated to MNE restructurings. There are several websites and databases available for information on restructuring cases. The European Restructuring Monitor (ERM) is an EU-funded publicly available website listing restructurings of enterprises announced in the media, with a focus on employment effects. The ERM offers a searchable database of restructuring events [see EMCC] based on announcements in national media sources. Private providers offer comprehensive information on mergers and acquisitions often provided in searchable databases.

For more information on globalisation in business statistics and macroeconomic statistics see [Eurostat] and [Eurostat 2019].

9. Conclusions

The EWS is intended as a clearly structured, light (non-legislative) procedure based on the voluntary cooperation between national data compilers and Eurostat for the recording of globalisation events in the European statistics. It has been set up to ensure consistency in terms of applied methods, statistical treatment and communication of results across the ESS and the ESCB.

It fully relies on the network of national EWS correspondents, coordinated by Eurostat. In order to keep the network up to date on the developments of the EWS, Eurostat organises regular webinars for the national EWS correspondents and ensures the management of the secure EWS collaborative platform.

A number of cases have been discussed within the EWS and so far, the chosen set-up has been adequate. The discussions on recent EWS cases concerning financial MNEs have shown the beneficial effects of full cooperation between the ESS and the ESCB.

The EWS is a learning process that is continuously improved as experience is acquired within the European context. The idea of going beyond European boarders and covering cases where countries outside the EU are involved in a globalisation event is currently under discussion.

References

1. CIRCABC (EC), Communication and Information Resource Centre for Administrations,
2. Businesses and Citizens, European Commission. CIRCABC 4.0.
3. EMCC, The European Monitoring Centre on Change restructuring events database.
4. Eurostat, Economic globalisation dedicated section for providing information on business statistics and economic statistics.
5. Eurostat (2017), Explanatory methodological note for the ESTAT news release on retail trade. Published on the Eurostat website, 4 May 2017.
6. Eurostat (2019), Globalisation – Overview of European ongoing activities and prioritisation. Paper prepared for the UNECE meeting on the Expert Group on National Accounts: Measuring Global Production, Geneva 10-12 April 2019, UNECE/CES/GE.20/2019/11.
7. Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics. OJ L 87, 31.3.2009, p. 164-173.

How do national targets for children measure up to global vision of the 2030 Agenda?

Renee Sorchik, Yanhong Zhang, Mark Hereward
UNICEF HQ, New York, USA

Abstract
Issues of national SDG target-setting and development of robust indicator frameworks have thus far received inadequate attention in global policy dialogues on SDG implementation. This study analysed the adoption, target-setting and level of ambition, and use of global indicators to measure progress among child-related targets and indicators in 93 middle and low income countries. Results show that while countries have adopted a significant share of the child-related targets, they have not yet set national target values. Further, there was no designated indicator to measure progress against targets incorporated into national development strategies in 36 percent of targets across the 93 countries. Supporting countries' national target setting and monitoring will require further attention if the 2030 goals are to be reached.

Keywords
SDG; Target-setting; Nationalization; Indicator; Data

1. Introduction

The 2030 Agenda emphasizes the centrality of nationally owned sustainable development strategies, as well as the primary responsibility that governments have for follow-up and review to support the implementation of the Agenda. In contrast to their importance, the issues of national SDG target-setting and development of robust indicator frameworks have thus far received inadequate attention in global policy dialogues on SDG implementation. For instance, while country reports submitted to the last three cycles of the High Level Political Forum on Sustainable Development (a platform to review SDG implementation at the global level as mandated by the United Nations General Assembly) have seen a noticeable increase in the use of indicator data to provide baseline of national SDG implementation, achievements, and gaps across the various goal areas, it is not clear whether the baselines and assessments were benchmarked against nationally set targets or if such targets existed at all.

The study presented in this paper is expected to fill an important knowledge gap by conducting a systematic mapping of national SDG target-setting and the establishment an indicator framework to monitor progress.

The findings are expected to inform debates at various levels on policy priority-setting in national SDG implementation. The findings should also shape discussions on where and how to optimize support to national statistical capacity development in order to fill the gaps in data quality and availability to meet the needs for monitoring.

2. Methodology

All 130 UNICEF Country Offices (COs) were requested to complete an Excel data collection tool enquiring about the country-specific situation of 23 SDG targets considered to directly improve the status of children and women, and 33 indicators from the SDG global monitoring framework measuring the progress towards the targets. As UNICEF COs are not present in high income countries, responses represent middle and low income countries.

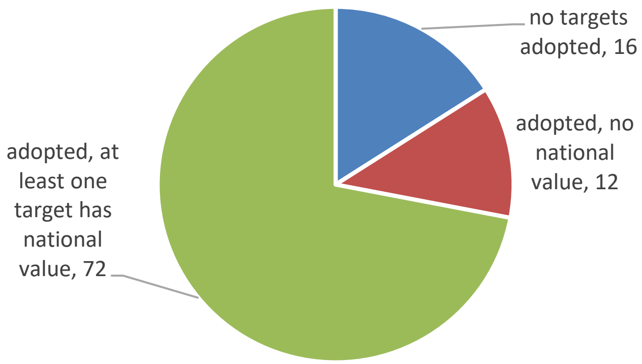
For the 23 SDG targets examined, COs were asked if the targets were adopted into national and sectoral strategies, and if adopted, if national target values had been set. COs were also asked if targets were adapted or altered to fit the country-specific context. If the target was incorporated into a national development strategy, COs were asked to indicate if: the global SDG indicator was used to measure progress against the target; if a country-specific indicator had been developed; or if an indicator framework had not yet been developed to monitor progress against the targets.

3. Results

The SDG Agenda developed targets, most of which have a pre-defined target value aiming to achieve a certain objective by 2030. We first sought to understand how many of the child-related targets were being adopted by countries, and how advanced countries were in their national target-setting process. A target was considered adopted when it was incorporated into a national or sectoral strategy, or when COs indicated it had been chosen as a priority target by the government. Target values did not need to be predefined in order for a target to be considered adopted by a country. However, in line with their commitment to the 2030 Agenda, countries are meant to set national target values for each target they adopt. As part of the nationalization process, countries have the option to set their own country-specific target values in line with the SDG global value, or choose a more or less ambitious target value in accordance with their country-specific circumstances. National target values were considered set when they were documented in either a national or sectoral strategy or other public government document.

Responses were received from UNICEF COs in 93 countries. Among the 93 countries examined, 72 percent had adopted and set national target values for at least one of the 23 child-related targets. This implies that almost three-quarters of countries have initiated political discussions and have identified at least an initial set of priority targets to focus on in their development agendas. Another 12 percent of countries adopted one or more of the targets, but have not yet set any national target values. These countries have identified at least an initial set of priority targets, but are in the beginning stages of the nationalization process. The remaining 16 percent of countries have not yet adopted any of the child-related targets examined. In most of these countries, COs indicated the prioritization and target-setting process had not been completed and in many cases it was stalled due to a change in government (Figure 1).

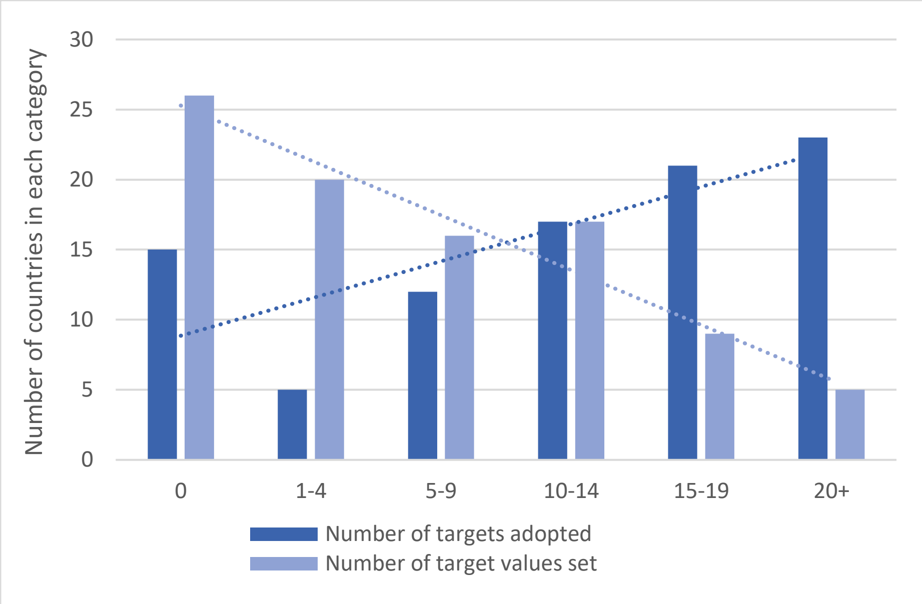
Figure 1: Status of adoption and nationalization of child-related targets



National target-setting

Figure 2 shows the relationships between the number of targets adopted and the number of national target values set. While almost half of countries adopted 15 or more of the child-related targets, indicating an interest in bettering the situation of children through the SDG Agenda, many of these countries have yet to set national values for these targets. If countries were further along in the national target-setting process, we would expect the distribution of the number of target values set to mirror that of the number of targets adopted. However, there is a negative relationship, demonstrating that while countries have adopted a fair number of child-related targets, they have not yet set national values for these targets. Just 14 countries set national target values for 15 or more targets, compared to 44 countries that adopted 15 or more targets. About half of countries set national target values for less than 5 targets, with 26 not having set any national target values. Among countries that have not set any national target values, 11 have adopted targets but not yet set any national values, and another 15 have not yet adopted any targets.

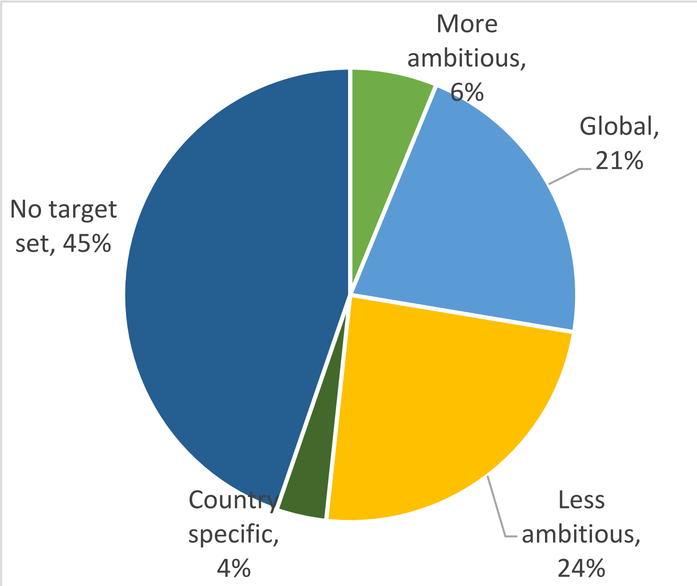
Figure 2: Number of countries by number of targets adopted, and by number of national target values set, for child-related targets



Target ambition level

Ambition levels of national target values were compared against global SDG values among countries that adopted targets. However, as many national target values have not yet been set (45 percent of targets across 93 countries, hereafter "country-target pairs"), relative ambition levels cannot accurately be determined (Figure 3). Countries may have politically adopted the target, but have not yet set a 2030 or other future aim for its achievement in almost half of the countries examined. In almost one in four country-target pairs, the national target value was less ambitious than the global value, and in about one in five country-target pairs the national target value equaled the global value. In just 6 percent of country-target pairs, the national target value was more ambitious than the global value. There were about 4 percent of country target pairs where the target values were country-specific and could not accurately be compared to global target values (Figure 3).

Figure 3: National target values in comparison to the global SDG values for adopted child-related targets



Use of global and country-specific indicators to monitor targets Global SDG indicators were developed to monitor targets and measure progress towards achieving the 2030 Agenda. These indicators were derived to be technically sound and rigorous, with metadata and technical guidelines outlining data collection, processing and analysis. Global use of these indicators would harmonize national data reporting as well as allow clear comparisons between countries and over time. Therefore, we were also interested in the proportion of countries using the global indicator to monitor progress of child-related targets, as well as how many countries created a country-specific indicator in lieu of the global indicator.

Generally speaking, if the parent target was incorporated into a national development strategy, the global indicator was used to measure progress of the target about half of the time (Figure 4). Country-specific indicators were used in lieu of global indicators in 15 percent of country-target pairs, meaning that further data collection or calculation would be necessary to make the indicator directly comparable to the global indicator. For 36 percent of country-target pairs, no indicator had been selected to monitor progress.

Therefore, only in about half of all country-target pairs, could the indicator be compared to regional or global averages, or be used to make inter-country comparisons or comparisons over time. Further data collection or analysis would be needed in order to report on the global SDG indicator in almost half of all countries examined.

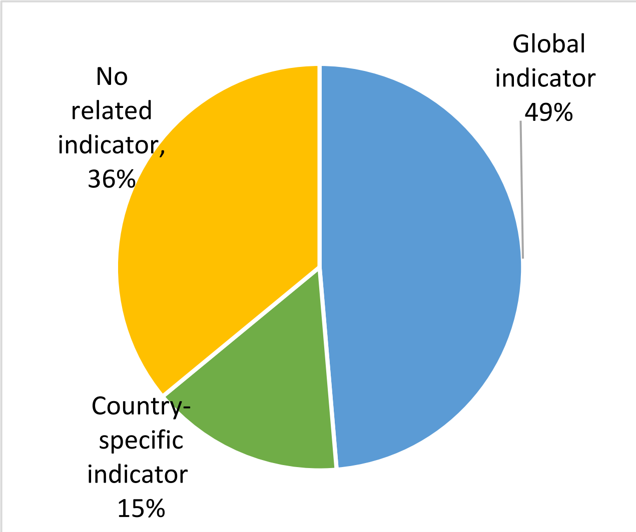
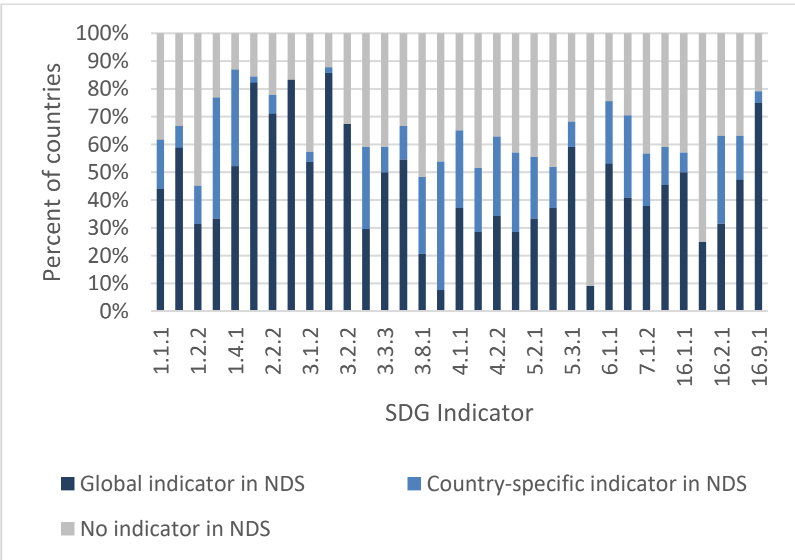


Figure 4: Percent distribution of indicators used to measure progress of targets in national strategic development plans

Figure 5 shows that use of the global indicator was not homogenous among the child-related indicators used to monitor progress of targets incorporated into national development strategies. Use of the global indicator ranged from 86 percent for indicator 3.2.1 (neonatal and under five mortality rates), to just 8 percent for indicator 3.B.1 (which was previously defined as the proportion of the population with access to affordable medicines and vaccines, but has more recently been altered to the proportion of the target population covered by all vaccines included in their national programme).

Figure 5: Percent distribution of global SDG, country-specific, or no indicator to monitor child-related targets incorporated into national development strategies



4. Discussion and Conclusion

While about half of the countries examined have adopted 15 or more of the 23 child-related targets, and 84 percent have adopted at least one of the priority targets, many countries have not yet set national target values. The prevalence of SDG monitoring and evaluation publications and voluntary national reviews may suggest countries are embracing the 2030 agenda and are full-heartedly tackling the world's greatest challenges, but 'you can't manage what you can't measure' and there is no way to measure progress against a target with no aim. Furthermore, in slightly more than one-third of country-target pairs examined (36%), there was no designated indicator to measure progress against targets incorporated into national development strategies. Even if a national target value had been set, without an indicator to measure the target, it is not possible to monitor country progress.

If this sub-set of targets can be used as a proxy for the larger SDG agenda, it would appear that middle- and low-income countries are not generally setting national target values that are more ambitious than the global values. However, as 45 percent of national target values still remain to be set, the ambition level of countries could drastically change going forward.

The results of this study suggest there is still work to be done for partner agencies to support countries in setting national target values and building robust monitoring frameworks to measure progress against national targets. Countries will need support to increase national statistical capacity in order to fill the gaps in data quality and availability to meet the needs for monitoring. Supporting national target setting and monitoring will require further attention and support if the 2030 goals are to be reached.

Measuring Sustainable Development Goals Indicators (SDGs) in Egypt: Challenges, opportunities, progress and innovations

Waleed Abdelkhalik
Statistician at National Statistical Office, Cairo, Egypt

Abstract

In March 2015 at its forty-sixth session, the United Nations Statistical Commission created an Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs), which is composed of representatives from a regionally-balanced group of Member States and includes regional and international agencies as observers. The group also invites as other key stakeholders, such as civil society, academia and the private sector, to attend its meetings and to provide inputs during consultations. The IAEG-SDGs was tasked with providing a proposal for a global indicator framework to follow up and review of the 2030 Agenda. In this context, the Member States adopted the 2030 Agenda for Sustainable Development In September 2015; the overarching principle of the 2030 Agenda for Sustainable Development is that no one should be left behind. To achieve that there is issues should take in account by National Statistical Offices (NSO) such as " Data which is high quality, accessible, timely, reliable and disaggregated by income, gender, age, race, ethnicity, migration status, disability and geographic location and other characteristics relevant in national contexts To support implementation at all levels. Based on above and To meet the ambitions and demands of the 2030 Agenda, NSO in all over the world has started to review their statistical survey and their statistical programs, capabilities to monitor the challenges of measuring and follow-up SDGs, so This paper aim to discuss the efforts exerted by NSO in Egypt to assess the current status for SDGs, and the challenges of assessing and monitoring SDGs, In fact, indicators, methodologies, innovations are crucial to face with these challenges and could be considered from different points of view. Challenges are opportunities for fruitful improvement of the statistical system in Egypt to fill data gap that related with SDGs

Keywords

Data Ecosystem; Egypt's Vision 2030; Data Disaggregation; SDGs report; the 2030 Agenda: SDG Observatory

1. Introduction

On 25 September 2015, world leaders meeting in New York adopted United Nations Resolution 70/1, "Transforming our World: the 2030 Agenda for Sustainable Development". The Agenda is "a plan of action for people, planet and prosperity", and "also seeks to strengthen universal peace in larger freedom".

The core of the Agenda is a set of 17 Sustainable Development Goals comprising 169 targets that draw on a large number of previous international agreements, especially concerning development, the environment and human rights.

The Goals are presented as "integrated and indivisible, global in nature and universally applicable". The Agenda presents them as addressing the 5Ps: People (broadly corresponding to Goals 1-5), Planet (6, 12, 13, 14 and 15), Prosperity (7-11), Peace (16) and Partnership (17).

Still, this does not mean that every target applies to every country. Rather: "Targets are defined as aspirational and global, with each Government setting its own national targets guided by the global level of ambition but taking into account national circumstances. Each Government will also decide how these aspirational and global targets should be incorporated into national planning processes, policies and strategies.

Based on above this paper aim to discuss the statistical efforts of monitoring and following up SDGs in Egypt(Challenges and opportunities).

2. Efforts for Monitoring and follow up SDGs in Egypt

1. Before the launch of the SDGs in September 2015, Egypt committed itself to achieving sustainable development in the Egyptian constitution (adopted in January 2014), The constitution covers the three dimensions of sustainable, it contains articles on strengthening governance, equality, and social justice. It is a modern constitution that ensures the constitutional rights of citizens, while guaranteeing that no one is left behind.
2. Egypt's Vision 2030 (SDS) was launched during the "Youth Year", on February 24, 2016, is comprehensive and multidimensional, and ensures the needs of current and future generations, also it serves as the national umbrella through which the SDGs will be implemented in Egypt. It reflects the priorities of Egyptians for a better future, the strategy will be implemented by all stakeholders, including the Government, the private sector, civil society organizations, etc, in a cohesive approach.
3. Establishing Sustainable Development Unit (SDU): According to the decision No. 18 of 2016 issued by CAPMAS' president, a working group was formed as the "Sustainable Development Unit" to do the following points:
   3.1 Coordinating between the different statistical departments inside NSO (Members of SDU) to obtain data for SDGs
   3.2 Identify the most important challenges that face monitoring SDGs, and build partnerships for this objective

3.3 Identify the most important capacity-building programs to meet the needs of SDGs.
3.4 Building an integrated database and update it Periodically to serve monitoring and follow up SDGs.
3.5 Preparation of statistical reports to monitor and follow up SDGs.
3.6 Raising awareness with SDG (posters - Info-graphic – social media - stories for children …etc).
3.7 Develop a action plan for preservation and documentation metadata related to SDG such as the outputs of any participation for members from NSO and SDU at national or international events also National SDG reports – database – document …etc) to build institutional memory and Electronic library.

4. Participation in the Inter-agency Expert Group on Sustainable Development Goals

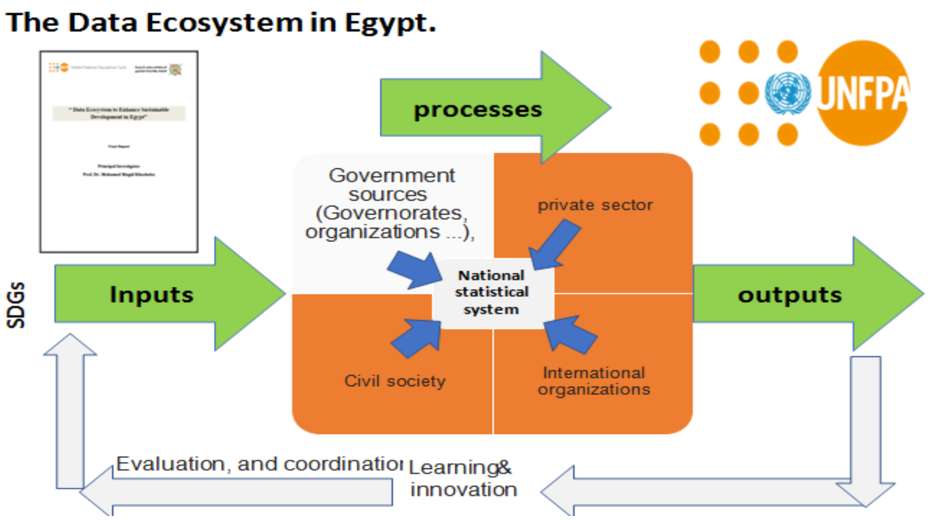
On 6th March 2015, at its forty-sixth session, the United Nations Statistical Commission created the Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs), composed of Member States and including regional and international agencies as observers. The IAEG-SDGs was tasked to develop and implement the global indicator framework for the Goals and targets of the 2030 Agenda

- The framework of SDGs is updated according to (Tier1-Tier2-Tier3) classifications, which classify indicators according to the clarity of concepts and the availability of methodologies and periodicity in production. In addition to communicating with the custodian agency to discuss updates in methodologies and concepts.
- Egypt is represented in the (IAEG-SDGs), which consists of 28 members and representatives of regional and international committees and agencies as observers.

5. Lunching Data Ecosystem report to Enhance Sustainable Development in Egypt

The report presents a future vision of the Data Ecosystem in Egypt, and the role of CAPMAS and new data communities in this system, in the light of a conceptual framework and acquired experiences as for the nature and execution of world initiatives to activate Data Revolution and Big Data, so as to enhance the objectives of sustainable development; characterize the components of the National Data Ecosystem, evaluate the status quo of data, shed light on new data communities in Egypt, analyse challenges facing system development, and crystallize a strategic vision to build and develop capacities of the system.

Figure 1: The Main Elements of Data Ecosystem in Egypt



Source: Report of Data Ecosystem to Enhance Sustainable Development in Egypt",2018.

6. **Studying governorates and localities' role in providing data to monitoring SDGs in Egypt**

The national statistical office was conducted field visits to 6 governorates (Menoufia, Ismailia, Alexandria, Minya, Beni Suef, El Behera) to identify the available data in the governorates and municipalities that can be used in monitoring SDGS, also determining the most important statistical capacity building programs needed by the governorates and municipalities, which help in monitoring SDGs' indicators.

7. **Coordinating with statistical partners to fill in the data gaps in SDGs' indicators**

For example: benefit from The UN Water Commission's initiatives to monitor and follow up Goal 6, "Ensuring the data availability and sustainable management of water and sanitation for all" to fill data gaps for the following indicators:

- ✓ 6.1.1: Proportion of population benefiting from safe drinking water services.
- ✓ 6.2.1: Percentage of population benefiting from the proper management of sanitation services, including hand washing facilities with water and soap.
- ✓ 6.3.1: Quantity of treated wastewater by type of treatment.
- ✓ 6.4.1: Change in water use efficiency over time.
- ✓ 6.5.1 Degree of implementing IWRM (0-100).

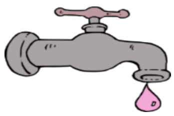
8. **Issuing the statistical report to monitor and follow up SDGs in Egypt.**

The national statistical report on SDGs' 2030 presents the current status of the indicators for which data are available and Highlighting some gaps and challenges related to some indicators that require further study so that they can be obtained in a accurate, systematic and comparable manner. (for more details see: National statistical report for SDG,2018 that available at CAPMAS website:www.capmas.gov.eg).

✓ **An Examples of SDG indicators that available in Egypt**

Goal 1: No Poverty - End poverty in all its forms everywhere.

i. Proportion of population living in households with access to basic services.



Connected to the main water Network 97%.- 2017

Connected to a safe sewerage system 66% -

Source: CAPMAS, 2017

Goal 2: End hunger, achieve food security and improved nutrition and sustainable agriculture.

2.2.1 Prevalence of stunting among children under 5 years of age (height for age <-2 standard deviation from the median by World Health Organization (WHO) Child Growth Standards)

The rate of stunting among children under the age of 5 years (21.4%).

Goal 3: Ensure healthy lives and promote well-being for all at all ages

3.1.1 Maternal mortality ratio

49 cases / 100,000 live births.
Source: Ministry of Health, 2015

3.1.2 Proportion of births attended by skilled health personnel

91.5%
Source: Ministry of Health, 2015

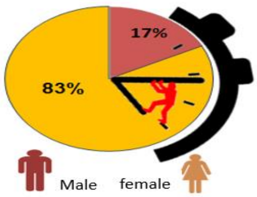
3.4.1 Mortality rate attributed to cardiovascular disease, cancer, diabetes or chronic respiratory disease .

36.4 Per 100,000 population	304.4 Per 100,000 population	9.5 Per 100,000 population	27.8 Per 100,000 population

Source: Ministry of Health, 2015

<u>Goal 5: Achieve gender equality and empower all women and girls.</u>

5.4.1 Proportion of time spent on unpaid work in domestic and caring activities, by sex, age and place of residence



The percentage of time spent in unpaid work for domestic and caring activities was 17% for males and 83% for females
Source: CAPMAS, 2015

<u>Goal 6: Ensure availability and sustainable management of water and sanitation for all</u>

6-5-1 Degree of implementation of integrated water resource management (IWRM (0-100) **.**

The Degree of implementing integrated water resource management (40%) Source: National statistical report for SDG,2018.

9. **Lunching the Egypt SDG Observatory for raising statistical awareness for all users**

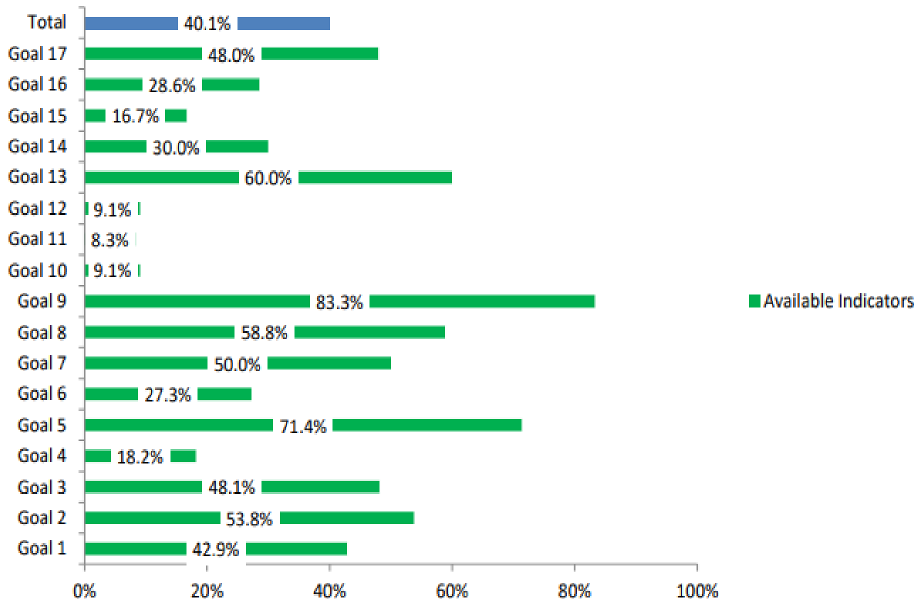
   The observatory provides a digital platform for tracking progress towards the SDGs in Egypt, also provides a summary of the number of indicators available, not available and not applicable for each goal.

   (See http://www.egyptsdgobservatory.info.)

3. The results of evaluation The Current Situation for SDGs in Egypt.

1. The first national statistical report on the SDG indicators was launched in May 2018; it outlines the classification of SDG indicators into the three tiers prescribed by the global indicator framework on the basis of their level of methodological development and the availability of data,(the percentage of indicator that Available was around 40%, Tier1(35,7%), Tier2(29.1%), Tier3(32.4%), not applicable (2.9%), notated that there are many indicators with gaps related with disaggregation such as migration ,disability …etc.)

Figure 2: the statistical situation of the SDGs in Egypt by Goal



Source: First National Statistical Report of SDGs in Egypt. Cairo: CAPMAS.2018.

2. The results of evaluations SDGs in Egypt refers to many challenges facing statistical coverage of goals and indicators of sustainable development 2030 in Egypt in the framework of data ecosystem, and classifies them into general, legislative, legal, organizational, technical, technological, human, and consultative challenges as follows:
   - The large increase in size and details of required data, and backwardness of tools and methods of collecting, processing, and analysing related Big Data.

- Absence of developing statistical legislations, and its reflection on statistical activity and relations with partners.
- Multiple challenges related to transforming administrative registers into statistical registers.
- Shortage of suitable finance and technical support, and developing national statistical cadres according to global standards.
- Gaps in data and methodologies, globally, regionally, and locally

4. Discussion and Conclusion (Challenges can be fruitful opportunities)

Measuring SDG indicators face significant challenges in Egypt, where the major challenge to SDG monitoring and follow up is expanding the goals as they cover a wide range of issues, including the human activity on earth, using water, energy, food, agriculture, health, sustainable consumption and production, manufacturing, urbanization, education, inequality, poverty, as well as gender issues. and disaggregated data that not available to ensure measuring the indicators and making international comparisons, however there are many opportunity to face this challenges by:

1. Speed up the finalization of the National Strategy for the Development of the Egyptian Statistical System, which will ensure the provision of a framework for partnership and collaboration with all the elements of the statistical system in a manner that serves the acceleration and legalize the processes of monitoring 2030 SDG indicators.
2. Promote communication with the partners producing data on the international/ regional/ national levels and in coordination and collaboration with NSO, in order to hold regular meetings to identify the most significant challenges, conduct technical reviews for the work methodologies, and determine the most important indicators that may need surveys, which may positively impact monitoring the progress towards SDGs.
3. Expand the use of the administrative records to produce official statistics and achieve the sustainable measurement of 2030 SDG indicators.
4. Introduce new sources and techniques to obtain the data related to sustainable development indicators and employ modern technologies, like using GIS and big data.
5. Strengthen the relationship between CAPMAS and all its partners in the statistical work by fostering the culture of participation in producing the indicators (the private sector, NGOs, scientific research centres, universities and others).

6. The need to disseminate the culture of paying attention to the data and information and raising awareness of their importance in planning, and accordingly the need to raise awareness of the 2030 SDGs.

References

1. Central Agency for Public Mobilization and Statistics (2018): The National Statistical Report of SDGs in Egypt.that available at http://www.egyptsdgobservatory.info
2. Central Agency for Public Mobilization and Statistics (2018): Report of Data Ecosystem to Enhance Sustainable Development in Egypt", Cairo, Egypt (unpublished report).
3. Waleed Mohammed (2017): The seven meeting of IAEG-SDG , presentation available at: https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-07/6.1%20Egypt.pdf
4. Sustainable Development Solutions Network (SDSN) (2015). "Indicators and a Monitoring framework for the sustainable Development Goals", Sustainable Development Solutions Network, United Nations, December 2015.
5. United Nations, (2011), "Sustainable Development in the 21st Century: implementation of Agenda 21and the Rio principles, Department of Economic and Social Affairs.
6. United Nations **(2016).** Report of the Inter-agency and Expert Group on Sustainable Development Indicators to the Statistical Commission, E/CN.3/2016/2/Rev.1.
7. Grosskurth, J. &J. Rotmans (2005). The Scene Model: Getting Grip on Sustainable Development in Policy Making. Environment, Development and Sustainability, 7, no.1, 2005,135–151.
8. Egypt's voluntary national review 2018 at the annual High-level Political Forum (HLPF) on 9-18 July, 2018 ,available at : https://sustainabledevelopment.un.org/memberstates/egypt
9. National Statistical System Assessment report (2015). EGYPT, September ,available at http://www.paris21.org/sites/default/files/NSS-Assessment-Egypt_2015.pdf.

Linking the Population and Housing Census with the Agricultural Census: Technological advances and new opportunities for developing countries

Aberash Tariku Abaye
Central Statistical Agency of Ethiopia (CSA)

Abstract

The World Census of Agriculture (WCA) 2010 program introduces modular census approach for conducting a census of agriculture. The WCA 2020 program clarifies the core and supplementary module. The core module is expected to be conducted on complete enumeration basis while supplementary modules to be conducted on sample basis. These modules should be conducted in short period of time 1 to 2 years. Data from the core module should be used as frame for the supplementary module. It is recommended that the core module include frame items to be used by supplementary module. The Food and Agriculture Organization (FAO) recommends countries to link the population census with agricultural census by including agricultural section in their population and housing census. The content may vary from country to country. Maps are used in preparing the frame, for conducting the field data collection and presenting and analyzing results. Preparation of suitable frame is essential for conducting agriculture census which can be used as master sampling frame for subsequent agricultural surveys. The list frame, area frame or multiple frame can be used. The experience in Ethiopia shows that the core module which is collected during population and housing census are number of agricultural holders and type of agriculture which is used to prepare frame for the annual agricultural survey. New samples for the annual agricultural survey are selected in 2018 based on the cartographic frame prepared for the 4th population and housing census. This survey is expected to provide more reliable and up-to-date data as we used the new updated frame. The enumerators also used electronic map for delineating and listing. This technological improvement saves time and improves quality.

Keywords

Modular Census; core module; supplementary module; population census; agriculture census

1. Introduction

The World Census of Agriculture (WCA) 2010 program introduces modular census approach for conducting a census of agriculture. The WCA 2020 program clarifies the core and supplementary module. The core module is expected to be conducted on complete enumeration basis while supplementary modules to be conducted on sample basis. These modules should be conducted in short period of time 1 to 2 years. Data from the core module should be used as frame for the supplementary module. It is recommended that the core module include frame items to be used by supplementary module. The Food and Agriculture Organization (FAO) recommends countries to link the population census with agricultural census by including agricultural section in their population and housing census. The content may vary from country to country.

Maps are used in preparing the frame, for conducting the field data collection and presenting and analyzing results. Preparation of suitable frame from population and housing census is essential for conducting agriculture census which can be used as master sampling frame for subsequent agricultural surveys. The list frame, area frame or multiple frame can be used.

2. Methodology and result

The Central Statistical Agency (CSA) is planning to conduct the 4th Population and Housing census. In preparation for this Census, CSA have been conducting the cartographic work over the past two years. Mobile devises are used for cartographic work. Location data for major features and each household is taken and geo referenced data base created. This helps CSA to have electronic copy of enumeration area maps with wel-defined boundaries. This frame of EAs provides an ideal frame for selecting sample for the household surveys. The CSA developed an updated EA summary database with records for all the census EAs that includes the geographic identification of each EA and the total number of households listed in each EA. This database is used for examining the distribution of the rural households by region and zone for the Annual Agricultural Sample Survey (AgSS) frame and to select the sample.

The geographic and administrative structure of Ethiopia is divided hierarchically into Regions, zones, weredas and kebeles. For the purposes of the census operations, each kebele is further divided into enumeration areas (EAs) designed to have about 150 to 200 households each in the urban areas and 100 to 150 households each in the rural areas. The EAs are used as primary sampling units (PSUs) for national household surveys. The EA map which is prepared during cartographic work will help to delineate the EA during the census and survey. The enumerator can use uses both electronic copy and hard copy of the map for delineation. For the 4th population and housing census 180,000 tablets, 126,000 power banks, 35,000 solar are prepared and around 152,374 EAs will be covered.

The experience in Ethiopia shows that the core module which is collected during population and housing census are number of agricultural households and type of agriculture which is used to prepare frame for the annual agricultural survey. In the cartographic work of population and housing census, households are listed and the list is used as frame for the 2018/19 annual agricultural survey. The same information which will help to build frame will also be collected during listing of population and housing census.

The main objective of AgSS is to measure the area and production of crops and the production of livestock at the regional and zone levels. The AgSS is also used to study different characteristics of the agricultural households, such as the use of irrigation and other agricultural inputs.

A new sample for the annual agricultural survey is selected in 2018 based on the cartographic frame prepared for the 4th population and housing census. This survey provided up-to-date data as we used the new updated frame. The enumerators also used electronic map for delineating and listing. This technological improvement saves time and improves quality.

The measure of size for the selection of EAs is based on the number households in the updated frame for each rural EA. Given that a very high percent of rural households are involved in agriculture, the total number of households was used as the measure of size for the rural EAs. The sampling frame includes new zones that were split from the zones in the previous frame. The previous AgSSonly included 66 zones; 10 zones in the Afar and Somali Regions with nomadic population were excluded. CSA covered all zones in the new frame for the AgSS, to cover all the crop area and livestock in Ethiopia.

The CSA has a frame of large commercial farms, which are covered in a special survey. This sector may be especially important for particular crops. This separate frame is important to ensure that the total crop and livestock production is covered since the large commercial farms are excluded from the AgSS,

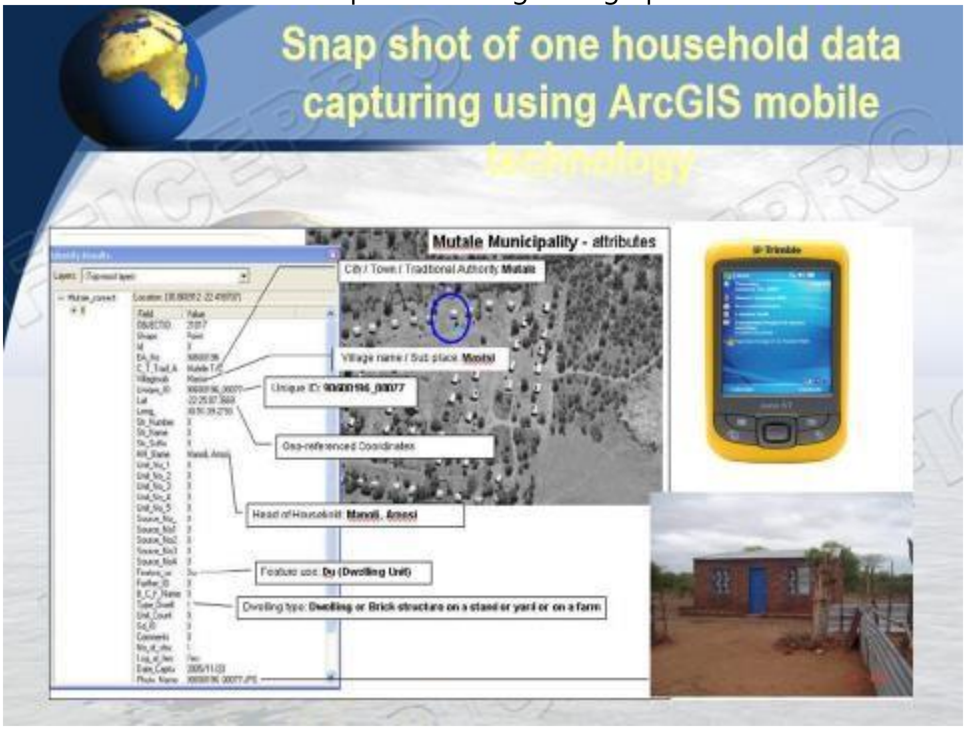
Getting reliable estimate for rare crops is a challenge in AgSS. Collecting information on the types of crops grown in each EA in the core module during population and housing census may help to develop a special frame for rare crops. The other two issues which is recommended to be covered in the core module are information related to irrigated crop area and livestock data on pastoral areas.

Technological advances save time and improve quality. The Enumeration Area map in the previous census was paper based. Identifying boundaries using paper based maps was difficult. There was also a problem in storage and identification of the maps for conducting sample surveys. The electronic EA maps prepared which are loaded on each tablet computer helps to guide the enumerator during delineation. . These electronic EA maps are also simple to access and print for other sample surveys.

3. Conclusion

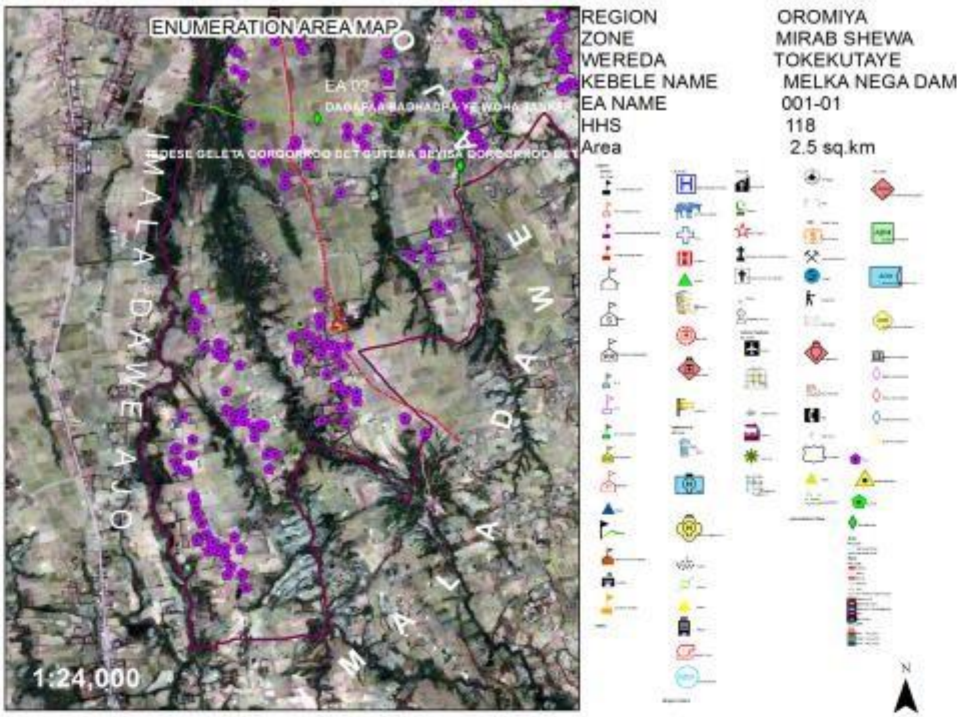
In conclusion Population and housing census should be utilized for collecting core information for agricultural census and survey. The experience from Ethiopia shows that electronic data collection during cartographic work and actual population census will improve quality of frame and data. It is also recommended that population and housing census should be utilized to get information for the rare crops, large scale farms and pastoral area data.

Annex I Household data captured during cartographic work

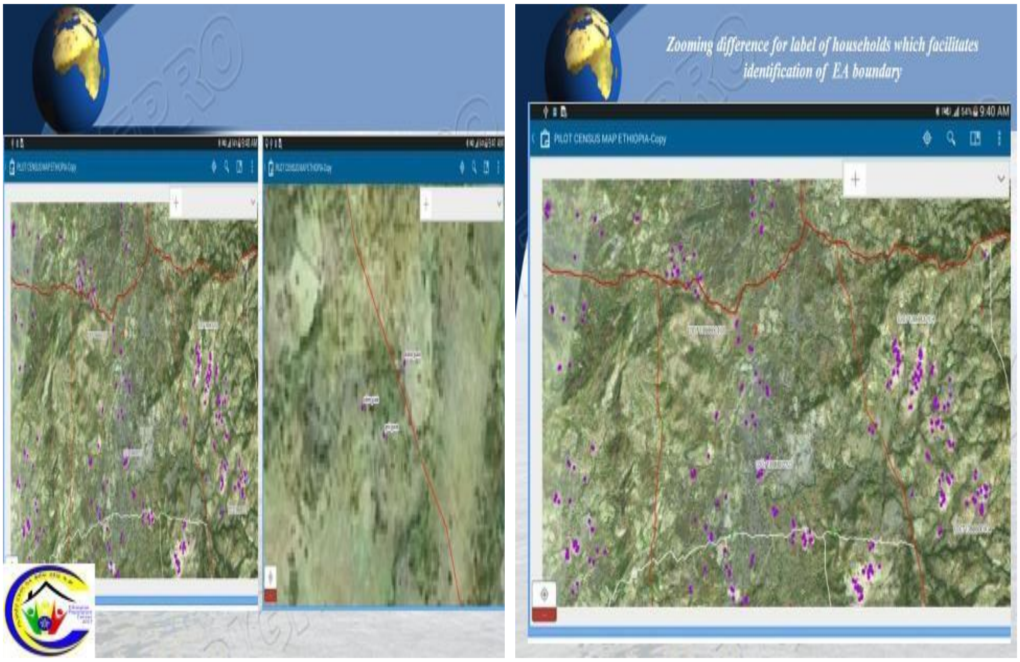


Annex II Use of map for field data collection

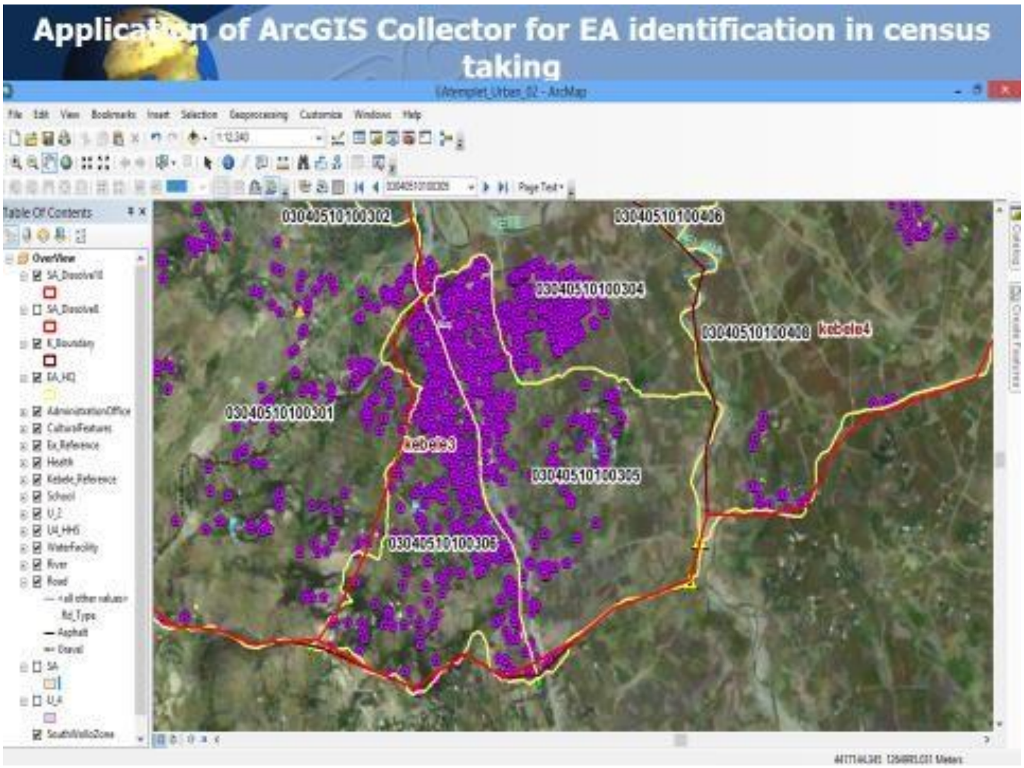
Annex III Enumeration area map



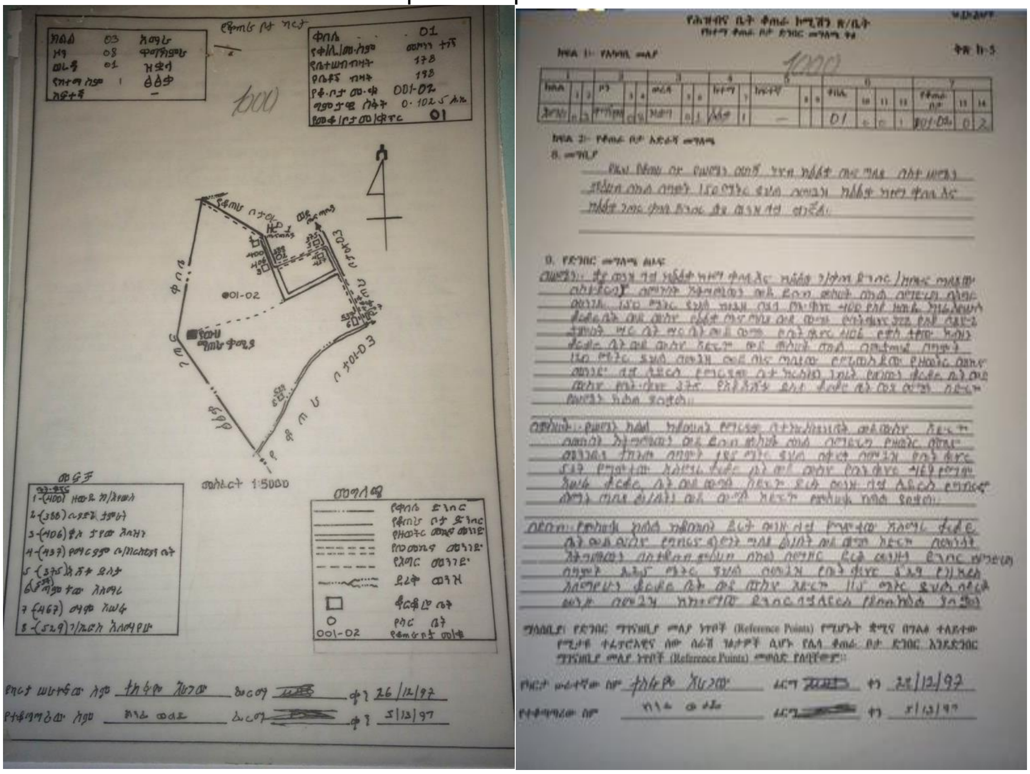
Annex IV Enumeration area map to be used by zooming



Annex V Enumeration Area map for EA identification



Annex VI Enumeration Area map in the previous censuses



Linking the Population and Housing Census with the Agricultural Census: Technological advances and new opportunities for developing countries (Georgian case study)

Maia Guntsadze
Tbilisi Open University

Abstract

This paper reviews innovative approaches used in Georgia for Linking the Population and Housing Census with the Agricultural Census. In November 2014 the National Statistics Office of Georgia (Geostat), for the first time, conducted Agricultural Census together with General Population Census. More precisely, there were four different censuses conducted at the same time: Housing Census, Population Census, Census of Emigrants and Agricultural Census. Aiming at preparing high-quality digital maps for census data collection/dissemination and using financial means/time in more effective way, Geostat implemented Geographic Information System (GIS). At the first stage, GIS group, which was created at Geostat, developed GIS maps using cartographic material available in the country. A supplementary work was done to cover the territories where cartographic material was missing. In order to minimize household/holding duplication and missing risks and effectively distribute work among census enumerators, Geostat conducted preliminary field works in 2013 and elaborated list of dwellings and households existing in Georgia. This information was digitalized and integrated to the cartographic data. Census of Agriculture for Legal Entities was conducted separately from General Population Census and covered all legal entities engaged in agriculture. It was the first time that Geostat used web-based questionnaire for Census of Agriculture and all enterprises participating in the census filled in these questionnaires. This paper will describe technological advances, new opportunities and benefits that Geostat received from conducting Census of Agriculture together with General Population Census, implementing GIS and using web-based questionnaires.

Keywords

Georgia; census; survey; statistics; agriculture; new opportunities

1. Introduction

Georgia is rich in agricultural tradition, which is an integral part of its history, mentality and cultural heritage. Agriculture played an important role in formation of the Georgian statehood and contributed much to its economic development. 43.4% (more than 3 million hectares) of the whole territory of Georgia is designated as agricultural land, which also includes pastures and meadows. 43% of the remaining area is covered with forest. Georgia has a wide variety of ecological and climatic zones conducive to the growth of for temperate climate and subtropical crops.

Total area of Georgia is 69,700 km2. 43.3% of the total area of the country (30.2 km2) is occupied by agricultural land (2005). Annual crops occupy 220.3 thousand hectares, and the perennial crops occupy 120.8 thousand hectares of land (2007). Besides, 43.2% of the country's territory is covered by forest fund (2010). If Georgia were an EU members state today, it would rank 17th based on its area, and it would represent 1.6% of the total area of the EU (EU28). Population of Georgia is 3,729.6 thousand people, 41.6% of which (1,554.8 thousand people) lives in rural areas (2018). 75% of the rural population is self-employed, mostly in the agricultural sector. Besides, in 2017, 43.1% of labor force was employed in agriculture, hunting and forestry, fishing and fishery industry, and total share of this sector was 8.2% in the GDP.

Almost all households living in villages are agricultural holdings and even in small towns 2/3 of households are engaged in agriculture. Overall, 3 out of 5 household are engaged in agriculture and they are spread through every region of the country. According to the Census of Agriculture 2014, total number of agricultural holdings is around 642 thousand, out of them only 2 thousand are legal entities while other 640 thousands are households. Majority of agricultural holdings are small and they produce agricultural products for their own consumption.

Due to favorable geographical location and climate, Georgia produces more than 25 kinds of permanent and more than 20 kinds of annual crops. Also animal husbandry is quite common in Georgia. One of the specificity of the country is that the agricultural sector is not well-specialized and majority of holdings produce many different kinds of agricultural products.

Statistical system in Georgia is centralized. The Law of Georgia on Official Statistics creates a general framework for statistical system in the country and sets coordination principles for all institutions responsible for producing official statistics in Georgia. According to the law, production and dissemination of statistics shall be based on the 10 basic UN principles of official statistics. Geostat is a coordinating body of Georgia's statistical system and the only producer of official agricultural statistics in the country[1]. Geostat is the agency responsible for conducting population and housing censuses, as well as agricultural census and economic censuses.

[1] The Georgian statistical system covers the whole area of the country, except the occupied territories

2. Methodology

In November 2014 the National Statistics Office of Georgia conducted Census of Agriculture together with the Population Census. In fact, there were 4 different censuses conducted together: Housing Census, Population Census, Census of Emigrants and Agricultural Census. Additionally, census identified holdings operating in the sector of aquaculture and greenhouses. The census was conducted using GIS maps.

Geostat started preparatory works for the General Population Census in 2006. One of the principal parts of these works was implementing GIS in order to create digital cartographic material and use them in census data collection and dissemination process. At the first stage, GIS group, which was created at Geostat, developed GIS maps using cartographic material available in the country. A supplementary work was done to cover the territories where cartographic material was missing. In addition to this, a numbering script was prepared in order to give its own code to each parcel and buildings located on it through the country territory. After this, a supplementary work was done for updating existing software and for switching to fully automatized system.

In 2013 Geostat conducted preliminary field works to elaborate list of dwellings and households existing in Georgia. In order to digitalize collected information and integrate it to the cartographic data, there was developed software which was using coding system created by GIS group. Information received from the preliminary field works was used for correcting and finalizing existing data on buildings and for distributing work among census enumerators effectively. In addition to these, household duplication and missing risks were minimized.

By implementing GIS, Geostat managed to significantly decrease time needed for preparing cartographic materials, use financial means in more effective way and prepare highquality maps.

Register of legal entities engaged in agriculture was prepared separately, due to the fact that General Population Census covered only household sector. To fulfill the list of agricultural enterprises, all existing reliable sources in the country were used: existing sample frame of the regular quarterly survey on agriculture, Statistical Business Register, Public register, Information available at the Ministry of Agriculture, at local governments, at the Ministry of Education and at the Patriarchate of Georgia. After the General Population Census Agricultural Census for Legal Entities was conducted separately.

Preparation process of the census questionnaires was held under the international guidelines and recommendations, taking into the consideration local user's needs. The census questionnaire consisted of four separate parts, respectively to four different censuses: Questionnaire on Dwelling (Q1), Personal Questionnaire (Q2), Questionnaire on Emigrants (Q3) and Questionnaire on Agriculture (Q4). In total, census questionnaire had 9 pages:

2 for Q1, 2 for Q2, 1 for Q3 and 4 for Q4. Each questionnaire had unique ID number which was linking them to each other.

Since the population and agricultural censuses were conducted together, it was agreed to use the same reporting unit, thus a household was considered as one agricultural holding. In case of two or more holders in the household one of them was selected as a holder and their operated lands and livestock were summed up and assigned to the holder. In case of one or more households in a holding, each household was considered as a separate holding and land and livestock were assigned proportionally according to the share of households in the holding.

Q1 contained questions on dwelling conditions, as well as the full list of households and persons living in the dwelling with status of the persons - usual resident, temporary absent, temporary present, emigrant. According to the status of the person, appropriate questionnaire was filled (Q2 or Q3). There were main identification questions in the Q1 for agricultural holdings. According to them, if household was identified as a holding, Q4 was filled by the enumerators. Besides, Q1 was identifying the holder. Additionally, Q1 contained identification question on holdings operating aquaculture. Q4 covered all core indicators recommended by the Food and Agriculture Organization and local user's needs were taking into account. Questionnaire on Agriculture (Q4) did not contain any question about the holder rather than holder's unique number indicated in Q1 and Q2, since all relevant information was covered by Q1 and Q2.

For the Agricultural Census for Legal Entities an electronic questionnaire was prepared, which was also used for the census of municipalities (common land data). The content of the electronic questionnaire was the same as of Q4. In addition to the questions similar to Q4, there were additional questions regarding personal characteristics of the company director/holder.

3. Results

As a result census database contains information on dwelling conditions, personal information of the population, personal information of emigrants, agricultural data on holdings, personal information of holders, data on households operating aquaculture and greenhouses and geographic data. The database is a comprehensive source for analysts and researchers to crossanalyze social-economic and agricultural data.

Coding system included in the census database easily links all obtained data and creates Master Sample Frame for Agriculture. Georgian MSF covers various variables regarding Family holdings, Agricultural enterprises, Households, Emigrants, Housing conditions, Aquaculture holdings, Greenhouse holdings, municipal (common) land, etc.

MSF facilitates identification and monitoring process in the field works, since it contains detailed personal data of the holder and her/his household members, as well as GPS coordinates of the household location. It also gives great possibility to prepare maps indicating sampled units for enumerators to be used during the field works. Thus, it will decrease the time needed for data collection and improve data quality.

Statistical Activities Following an Agricultural Census continues with dissemination of census data, including:

- Reconcile census data with current statistics
- Evaluate current system of statistics
- Develop or update list sampling frames
- Evaluate new methodology and technology
- Design new agricultural surveys or revise existing surveys
- Share experience conducting an agricultural census

Geostat intensively cooperates with international organizations and donors in order to implement the best practices and improve survey process. USDA has been helping Geostat since 2007 by providing trainings for permanent staff and enumerators, suggesting study tours and business trips and granting equipments. Additionally, USDA is shares expertise in developing methodologies and is organizes data users and producers meetings.

Based on Census results and in close cooperation with FAO, Geostat managed to:

- ✓ Prepare new sampling frame
- ✓ Create new sample (12 000)
- ✓ Update sample design and prepare rotational scheme
- ✓ Upgrade weighting methodology
- ✓ Develop estimation procedures
- ✓ Prepare recommendation for updating existing questionnaire design
- ✓ Elaborate mechanisms for automatic calculation of errors
- ✓ Create sampling frame update procedures

4. Discussion and Conclusion

As a result of census data and all abovementioned activities quality of Agricultural statistics increased dramatically in Georgia. In particular:

- ✓ Coverage of regional statistics increased
- ✓ Up to 150 regional figures released
- ✓ Up to 40 new indicators are be available
  - ➢ More detailed crops statistics
  - ➢ More detailed livestock statistics
  - ➢ More detailed farm-gate price statistics
  - ➢ New indicators on irrigation (irrigated areas)

  - ➢ Additional data on fertilizer use (types)
- ✓ New surveys implemented and data released
  - ➢ Pilot survey on slaughterhouses, elevators and refrigerators
  - ➢ Dissemination of the results of Census of Agriculture 2014
  - ➢ Quarterly survey on slaughterhouses, elevators and refrigerators
  - ➢ Survey on greenhouses
  - ➢ Survey on aquacultures
- ✓ Launched CAPI(World Bank-Survey Solution) for data collection in all Agricultural surveys
  - ➢ Non-sampling errors decreased
  - ➢ Timeliness and quality of data collection increased

Geostat works intensively with FAO experts. One of the important results of this mutual work is "Strategic Plan for Agricultural, Environmental and Rural Statistics (SPAERS)", which was elaborated by international and local Experts of FAO in close cooperation with Geostat and it has become the major roadmap for improving agricultural statistics through 2016-2020. The purpose of the documents is to improve the quality and coverage of agricultural and rural statistics and support the Government of Georgia in evidence-based decision making. SPAERS includes detailed timeframe of planned activities which are supposed to be completed in order to improve agricultural statistics in the country. Document also contains assessment and evaluation of current situation where it was highlighted that the unit cost of carrying out surveys in Georgia is considerably low compared to the countries with similar capacities. Now Georgia as one of pilot country is implementing the Agricultural Integrated Survey (AGRIS) funded by USAID, Bill and Melinda Gates Foundations (BMGF). The Main objectives are to optimize the scope of ongoing surveys based on the AGRIS methodology and rotating modules; prepare an action plan for calculation of priority SDG indicators.

Thus Linking the Population and Housing Census with the Agricultural Census was grasping opportunity to develop agricultural statistics in Georgia.

References

1. http://census.ge/en/results/census1
2. http://census.ge/en/methodology/kitkhvarebi
3. http://census.ge/en/methodology/categories
4. http://census.ge/files/results/agriculture/Agricultural_Census_2014.pdf

A European perspective to observing and reporting on SPEs

Tjeerd Jellema, Fausto Pastoris, Carmen Picon-Aguilar
European Central Bank

Abstract
The introduction of the new statistical standards incorporated additional guidance on the relevance of a comprehensive reporting on institutional units, and the identification of special purpose entities (SPEs), but stopped short of providing a unified definition based on clear criteria. The manuals did provide several characteristics, and these were most concretely formulated in the field of foreign direct investment, together with a request of the separate reporting of SPEs. Moreover, many countries noted the emergence of so-called near SPEs, e.g. corporations that would fulfil SPE like objectives whilst carrying out a measure of regular business functions. Existing statistical guidance allowed for certain flexibility in countries' responses, and thus led to a lack of comparability in individual countries' reporting. Efforts to introduce reporting requirements with respect to SPEs were hindered by the absence of a harmonized working definition of SPEs. The availability of a harmonized definition, decision tree and typology will assist compilers in Europe and world-wide to ensure completeness in reporting and provide separate reporting on SPEs in the context of external sector statistics. This paper reports on ongoing European efforts in the reporting of SPEs and the way forward, and will address the challenges faced by balance of payments compilers to achieve a consistent recording of SPE activity as well as the impact on key economic aggregates.

Keywords
SPE; Globalization; Balance of Payments; National Accounts

1. Introduction

SPEs are recognised, but not strictly defined, in the main methodological tomes dealing with macroeconomic statistics. They are described as being institutional units with no substance behind them, no physical presence, zero employment or physical output in terms of physical goods and services. Their presence in economic territory is largely a legal one, and their owners seek to realize certain advantages offered by the economic territory where they have legal presence. They may be also difficult to observe, and many countries struggle in correctly capturing transactions and positions of these entities. For countries that have concentrations of SPEs they may dominate the balance of payments financial account and the international investment position.

The paper is structured as follows. In a first section the paper reviews the characteristics of SPEs as captured by the IMF BOPCOM definition, and touches on the different types of entities that can be observed. The paper then explores the European dimension as regards SPEs, and explores some of the characteristics of the countries that are favourable to the establishment of SPEs and SPE like entities. Subsequently the paper then addresses the relevance of SPEs for euro area macro-economic statistics, notably the balance of payments / IIP, but also the financial accounts. Finally the paper concludes and describes the European current data collection effort.

2. Elements of the definition

The 2018 BOPCOM discussed and approved the definition of SPEs proposed by the IMF TF-SPE final report [IMF 2018], reproduced below. We briefly review these criteria, and qualify how these criteria should best be used. One aspect of the BOPCOM definition that should be mentioned is that it targets external sector statistics (ESS) explicitly, and thus that some of the criteria provided would be in need to be reviewed in the broader context of national accounts. In several aspects the definition breaks new grounds. First, in determining what SPEs are in the context of ESS all of the criteria in the definition need to be met. Thus the definition is necessarily more restrictive than the current looser formulations in BPM6 and SNA2008 that allowed compilers much more freedom to adjust to local circumstances, causing lack of comparability across countries.

IMF TF-SPE definition:

- An SPE resident in an economy, is a formally registered and/or incorporated legal entity recognized as an institutional unit, with no or little employment up to a maximum of five employees, no or little physical presence, and no or little physical production in the host economy.
- SPEs are directly or indirectly controlled by non-residents.
- SPEs are established to obtain specific advantages provided by the host jurisdiction with an objective to (i) grant its owner(s) access to capital markets or sophisticated financial services; and/or (ii) isolate owner(s) from financial risks; and/or (iii)reduce regulatory and tax burden; and/or (iv)safeguard confidentiality of their transactions and owner(s).

A key aspect of the definition follows from the observation that SPEs are by necessity legal entities registered/incorporated in the economy to be recognized as institutional units, yet do not have a physical presence to speak of, do not engage in physical production, and have no or little employment.

These elements of the definition are mutually reinforcing. Physical presence, output and employment are grounds for an entity to be considered a resident unit per sé.
The second element in the definition, states that the entity is under control by non-residents. A foreign parent is considered in BPM6 as a sufficient condition for an entity to be recognized as an institutional unit [IMF 2009]. This part of the definition does not preclude the possibility of resident direct or indirect control, if the entity would pass the institutional unit test in BPM6. IMF 2018 considered this possibility not relevant from the viewpoint of the ESS. The formulation used in IMF 2018 is borrowed from the FDI literature, extends the concept of a direct foreign parent to direct and indirect control (i.e. a majority vote in the shareholders meeting). It allows the FDI concept of subsidiary to match the SPE definition.

A novelty in the definition is that it integrates four sets of reasons why an entity can be considered to possess a 'special purpose', rather than referring to it in supporting text. This is also the most difficult part to operationalise, because it is as much a criterion describing the particularities of the jurisdiction of the host economy as it is an aspect of the entities themselves. These four reasons are helpful in the identification of SPEs by also scrutinizing whether specific host countries would fulfil the necessary conditions for hosting SPEs.

The final part of the definition again references relevance for ESS by formulating a requirement that a significant part of the balance sheet of the entity is vis-à-vis non-residents. This may prove to be too limiting and seems inspired by the FDI background of earlier work on identifying SPEs. It presupposes that the balance sheet consists mostly of financial instruments, specifically intra-group positions captured in FDI, however, is well know that SPEs also have assets or liabilities where the residence or the issuer/holder cannot a-priori be restricted to non-residents, such as tradable securities, and non-financial assets (intangibles, mobile equipment).

The IMF TF-SPE definition has unresolved issues, acknowledged in the report. Chief among whom is the treatment of entities that by convention or otherwise are recognised as institutional units irrespective of the economic territory of residence of the parent entity. For instance, legal entities established and controlled by households are considered separate institutional units from the household, up to and including household trusts. Likewise securitization vehicles (or special purpose vehicles) are not only considered specific financial instruments, but are also legal entities that are distinct from their nominal parent. They are considered separate institutional units classified as other financial intermediaries (S125).

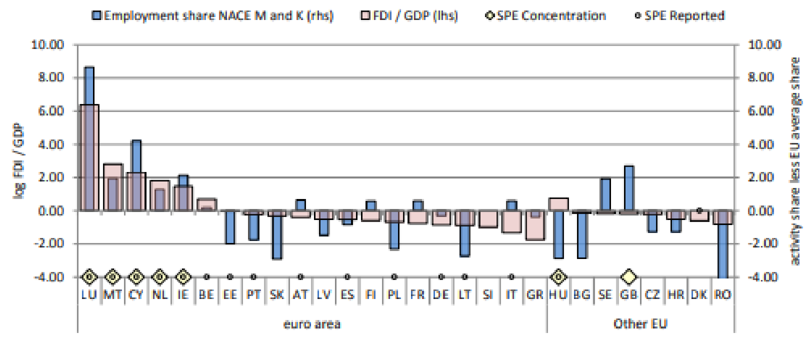
Finally, it may be necessary to draw a distinction between jurisdictions and economic territories. It may be the case, say in countries with a federal structure, or countries that have off shore locations as part of their economic territory, that there are differences between the various jurisdictions within an economic territory to create sufficient incentives for the creation of SPEs and SPE like entities. An example would be the island of Madeira that is part of the economic territory of Portugal, and a rather more prominent example would be the state of Delaware in the USA, which hosts very many of the financing and asset holding subsidiaries for large multinational corporations.

3. SPEs, country specialization and Europe

With pervasive globalization worldwide, implying a geographical specialisation of business functions related to the production and distribution of goods and services by large multinational corporations has been accompanied by a geographical specialisation of corporate financial functions. Off-shore jurisdictions are used by both corporations and private persons in order to optimize their tax liabilities. The European experience deviates. The economic weight of a single market and the attraction of a single currency provides to non-European multinationals ample incentive to establish subsidiaries there. A highly sophisticated financial sector is able to provide specialized financing products. Where national law prevails however countries have incentives to specialize to attract investors. MNE's likewise specialize by fragmenting their different business functions across EU member states. As a result the burden of statistically observing SPEs falls mostly on a limited number of European countries specialized in hosting them.

An ECB stocktaking on the presence and characteristics of SPEs by end-2016 [ECB (2017)] established that 16 out of the 22 responding EU countries report the presence of SPEs, and that 11 countries claim to comprehensively cover SPEs in the context of ESS. Most do so in the context of FDI, but a significant number also reports SPEs engaging in both, portfolio (PI) and other investment (OI) categories. Likewise, 9 countries report SPEs having non-financial assets on the balance sheet. As regards the sector classification, 14 countries report SPEs to be mostly classified as captive financial institutions (S127), yet countries also identify SPEs in the subsector other financial intermediation (S125), Insurance corporations (S128) and non-financial corporations (S11). There are several countries that are a priori suspected to have significant SPE populations, these are CY, IE, NL, MT, LU in the euro area and HU and GB in the other EU). In the remainder of this paper we refer to these as SPE concentrating countries. These countries typically show a larger financial balance vis-àvis the rest of the world in relation to GDP. One expression of this could be a FDI size indicator, as shown in Figure 1, which is calculated as the natural logarithm of the ratio of total FDI liability positions to GDP. Positive numbers indicate that FDI liabilities exceed GDP, this can be observed for seven countries. In Figure 1 the SPE concentrating countries are labelled by a diamond and countries that reported the presence of SPEs in [ECB (2017)] are labelled by a small circle.

Figure 1 Characteristics of countries with SPE concentrations



Source: ECB, Eurostat

In a European context, the establishment of SPEs structured to manage the tax liabilities and/or financing requirement of large MNE's requires highly specialised services. These are typically provided by legal professionals, tax consultants, specializing in international, European and home country law as well as financial sector specialists. Often these are provided by multinational consulting firms, with international staff, able to call on national expertise as and when needed. It is therefore of interest to consider whether SPE hosting countries would have a well-developed legal, financial and consulting services sector. Figure 1 shows the share of employment in financial and professional services as percentage of total employment. The indicator has been normalised against the EU average of 8.5%, countries with negative scores have less than the EU average, and countries with positive scores exceed the European average. All of the countries that we associate with concentrations of SPEs have positive scores exceeding 1, with the exception of Hungary. From the OECD restrictiveness indicator [OECD 2019a] we learn that European SPE concentrating countries have no restrictions with regards to FDI in the financial and professional sectors i.e. are fully open to international consulting firms.

One final and very compelling reason by which European countries may attract SPEs, and the first cited reason in the macro-economic statistics manuals [See for instance IMF 2009 para 4.50], is the provision of tax advantages. Often, these tax advantages are available to large corporations through the establishment of complex corporate legal structures in several countries in combination, rather than in single countries in isolation. Nonetheless, overall corporate income tax rates provide corporations with incentives to seek to generate corporate income in those countries where their tax liability would be minimised. The OECD tax database [OECD2019b] provides relevant data, comparing corporate income tax rates of individual

OECD countries, and tracing these over time. Table 1 summarises the available information for 2010 and 2018, with the same country grouping as presented earlier.

		2010	2018	Change 2010-2018	2018 Minimum	2018 Maximum
European Union	SPE Concentrating countries	22.3	18.8	-3.5	9.0	26.0
	Other	25.0	22.0	-3.0	15.0	34.4
OECD less EU	United States	39.2	25.8	-13.4		
	Japan	39.5	29.7	-9.8		
	Other OECD	25.0	25.0	0.0		

Source: OECD

The table conveys several messages. First, a number of countries have considered it necessary to significantly lower their corporate income tax rates since 2010, which will have a potential impact on corporate tax strategies. Not only the US with its recent tax reform, but also Japan, and European countries have significantly lower corporate tax rates in 2018 compared to 2010. Second, EU SPE concentrating countries differ strongly, some clearly adopted a low tax strategy (as have some nonSPE concentrating countries) whereas others are above median values of non-SPE concentrating countries. Headline income tax rates do not seem to be the unique responsible factor for concentrations of SPEs in Europe.

4. Measurement and quality

The ECB fact finding [ECB 2017] assessed how relevant SPEs were for the overall i.i.p. Three SPE concentrating countries responding to the survey (see Table 2 below) indicated that SPEs accounted for more than half of the i.i.p. total assets. Other responding countries listed between 0 and 10 percent. To a very large degree, the assets belong to the functional category of FDI, although portfolio and other investment, and financial derivative assets were also reported. Similarly, FDI was also prevalent on the liability side, but not so exclusively. For example the Netherlands and Luxembourg reported significant amounts as portfolio and other investment. There is thus a large impact of SPEs on the total i.i.p. of SPE concentrating countries and consequently the accurate and consistent recording of SPEs transactions and positions is highly relevant for the quality of the national b.o.p./i.i.p. as well as for European aggregates.

Such desired quality may however be difficult to achieve. Table 3 presents the intra euro area aggregate bilateral FDI asset positions as well as the mirror liability positions. Here euro area countries are classified as to whether they are considered SPE concentrating countries or not. The data are expressed as percentages of total intra euro area assets.

Total intra euro area liabilities are 91% of total assets therefore the intra euro area asymmetry amounts to 9% of total assets. Note that SPE concentrating countries hold 60% of the total intra euro area assets, of which 34% are liabilities vis-à-vis SPE concentrating countries, and 26% are vis-à-vis other euro area countries. The breakdown of the bilateral asymmetries is also relevant, 3% is between SPE concentrating countries, and 5% refers to mismatches between SPE concentrating and other euro area countries. Bilateral asymmetries between euro area non-SPE concentrating countries are only 1% of total intra euro area assets, These asymmetries and the overall relevance of SPE concentrating countries indicate the importance of SPEs for the quality of national as well as euro area aggregates. Some consideration must be given to the problem of pass through investment, which is a typical concern in the field of FDI statistics, and indeed has been one of the main drivers behind the separate reporting of SPEs to Eurostat and OECD. In recent years, the validity of the assumption that SPEs are predominantly responsible for pass through investment has been challenged, and alternative methods of measuring pass through investment have been proposed (Leino et al. [2015], Kocerka et al [2017]).

Table 2 ECB Fact finding on SPEs – relevance of SPEs positions by functional category

	Assets					Liabilities				
	Total	DI	PI	OI	FD	Total	DI	PI	OI	FD
AT	11%	32%				10%	33%			
BE	7%	16%				9%	18%			
ES	2%	5%	1%	0%	0%	2%	6%	1%	1%	0%
LT	0%	1%				0%	4%			
LU	53%	96%	6%	15%	9%	52%	95%	16%	25%	8%
NL	51%	77%	4%	12%	23%	55%	83%	27%	14%	11%
PT	5%	13%	0%	5%	0%	3%	11%	0%	0%	
CH	3%	8%	0%	4%	0%	4%	11%	0%	1%	0%
DK	2%	9%	1%	0%	0%	2%	12%	0%	0%	0%
HU	55%	64%	0%	1%	2%	37%	52%	0%	0%	0%
PL	1%	2%				0%	1%			
SE	4%	4%				7%	7%			

Source: ECB

Table 3 FDI Positions, reported assets and mirror liabilities, percentage of total intra euro area assets, 2017

			euro area		Intra EA Total Assets
			SPE concentrating countries	Sparse or No SPE countries	
euro area	SPE	Reported	34	26	60
		Mirror	31	24	
		Asym.	3	2	
	NSPE	Reported	22	18	40
		Mirror	19	17	
		Asym.	3	1	
Intra EA		Total Liabilities	49	42	100

% of total intra EA Assets

Source: ECB

It has been observed previously that SPEs may occur in different institutional subsectors, even including non-financial corporations. Two financial corporation subsectors, other financial intermediaries (S125) and captive financial institutions and money lenders (S127) are however considered particularly relevant in relation to SPEs. This was confirmed by the ECB fact finding [ECB (2017)]. The sector detail provided in the European b.o.p. does not allow a detailed assessment of the relative importance of these two subsectors by country. More sectoral detail is however available from the monetary union financial accounts (MUFA) that provides an aggregate for subsectors S125, financial auxiliaries (S126) and S127 that is denoted by the code S12O. As subsector S126 would lack a substantial balance sheet, the S12O balance sheet mostly reflects on the two sub-sectors of interest. In order to establish the relative importance of these sub-sectors as regards the presence of SPEs, we focus on their equity and debt securities liabilities. Whereas entities classified in S125 would have by and large liabilities consisting of debt securities but relatively little equity, entities belonging in S127 might have either equity or debt securities as liabilities. By assessing the relative size of liabilities of the combined sector S12O in the economy a sense of the importance of these types of SPEs can be obtained. We assess the share of equity liabilities for S12O which captures for instance holding companies (S127) compared with the total economy (S1). We also calculate the share of debt security liabilities in S12O to the total debt security liabilities of S11 and S12O (associated with either S125 or S127), as well as the share of debt security liabilities of securitization vehicles (SPV's, exclusively classified in S125).

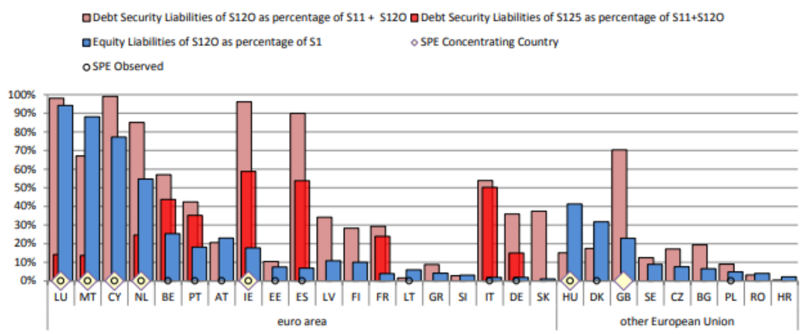
Figure 2 presents these data. The seven countries with assumed SPE concentrations are labelled with a yellow diamond. As regards the equity indicator, the euro area SPE concentrating countries all report high shares (between 55% for the Netherlands to 95% for Luxembourg, only Ireland has a relatively low 20%). As regards the debt securities indicator, values exceeding 50% are observed for all SPE concentrating countries (except Hungary). It is noteworthy however that high values are observed also for several non-SPE concentrating countries such as Spain, Italy and Germany. Part of the explanation is found in the relevance of SPV's (S125) in these euro area[1]

[1] The data is not available for other EU countries

The ECB has started recently a separate voluntary collection from EU countries of quarterly b.o.p. and i.i.p. data for resident SPEs. This collection covers some broad details of the current and financial account, a split by financial and non-financial SPEs and basic geography to understand the impact of SPEs in the EU. This collection is currently based on national definitions; however the idea is to move to the IMF TF-SPE definition to make this definition operational supported by the European typology.

countries. The difference between total S12O debt securities liabilities and SPV liabilities may be attributable to the existence of captive financial institutions (S127) that issue debt securities on behalf of related corporations, so-called conduits.

Figure 2 SPE visibility in the financial accounts



Source ECB

5. Conclusion

The work by the BOPCOM TF in providing a working definition of SPEs will benefit the comparability of external sector statistics. The working definition will benefit European and worldwide compilers in harmonizing their reporting on SPEs and allowing a consistent labelling of SPEs across Europe. The BOPCOM definition stops short of considering the case when entities are not directly or indirectly foreign controlled, pointing at the need to complementing the BOPCOM definition from the national accounts side.

The presence of SPEs can be circumstantially evidenced through assessing whether an infrastructure exists in a given country to support a significant SPE population. A clear entry point is knowledge of particular legal provisions in a specific jurisdiction that sets SPEs apart from regular corporations. At a first glance, the presence of legal and financial service sector with an employment share over the European average seems a good indicator, as would be the openness to investment abroad to large international legal, accounting and consulting firms. In a few SPE concentrating countries, low headline corporate tax rates would seem a locational factor as well.

In SPE concentrating countries, the impact on the national international investment position is large, and clearly motivates the separate accurate reporting of SPEs, as SPE related flows and positions would dominate the 'regular' flows. Moreover, it is clearly established that SPEs in the European context are relevant not only in FDI, but also with regards to other functional categories, notably portfolio investment, may have non-financial assets and can be found as non-financial corporations (S11) as well as financial corporations, in particular other financial intermediaries (S125), captive financial institutions (S127) and insurance corporations (S128).

References

1. IMF (2018) IMF BOPCOM Final Report of the Task Force on Special Purpose Entities (BOPCOM 18/03)
2. IMF (2009), Balance of payments and international investment position manual – 6th edition.
3. ECB (2017), Results of the fact-finding exercise on SPEs (mimeo)
4. ECB (2018), Defining a set of reporting requirements for Special Purpose Entities (mimeo)
5. OECD (2009), OECD Benchmark Definition of Foreign Direct Investment - 4th Edition
6. OECD (2019a) Corporate Tax Statistics Database
7. OECD (2019b) Regulatory Restrictiveness Indicator.
8. Leino, T. & Ali-Yrkkö, J. [2014]. How well does foreign direct investment measure real investments by foreign-owned companies? – Firm level analysis. Bank of Finland Research Discussion Papers 12/2014.
9. Kocerka, J, Makowski, K, [2017] 'Does pass-through funds require their own functional category?', ISI 2017.

The new international classification of status in employment and how it can be measured in a Labour Force Survey

Michael Frosch
ILO, Department of Statistics

Abstract

The aim of the paper is to outline the new Resolution concerning statics on work relationships adopted at the 20th International Conference of Labour Statisticians in October 2018 which replaces the standards adopted in 1993 known as International Classification of Status in Employment (ICSE-93). The paper highlights the main changes that have been introduced in the statistical framework regarding work relationships which will impact on the data collection of status in employment. In addition, the paper addresses the issue of how the data collection according to the Classification of Status in Employment 2018 (ICSE-18) can be structured in a labour force survey. The paper focuses on the measurement of ICSE-18 using a Labour Force Survey instrument with particular attention to the measurement of the new category of dependent contractors. ICSE-18 comprises ten mutually exclusive categories that can be organised according to type of authority and type of economic risk. In addition there is the extended classification of status at work which provides a categorisation of work relationships for all forms of work, including unpaid forms of work. The proposed measurement approach for measuring ICSE-18 builds upon the current widespread practice in household surveys of using one or two questions for collecting status in employment as defined in ICSE-93. However, this practice will not be sufficient to determine the status in employment as defined by ICSE-18. In addition, a set of classification and sub-classification modules are required that will either verify the respondents' initial response or re-classify the status in employment. In general, countries have little experience of how to measure the new category of dependent contractors contained in the statistical framework of status in employment. The identification of this group is a challenging aspect of the new framework. Different approaches to operationalization have been tested and some experience has been developed. This aspect together with other challenging issues of the new statistical framework of status in employment are highlighted in the paper.

Keywords

20th ICLS Resolution I concerning statistics on work relationships; International Classification of Status in Employment 2018 (ICSE-18); Labour Force Survey; Data collection

1. Introduction

The Resolution concerning statistics on work relationships adopted at the 20th International Conference of Labour Statisticians in 2018 (20th ICLS Resolution I) (ILO, 2018a), replaced the previous standards defined in the International Classification of Status in Employment (ICSE-93) (ILO, 1993). The 20th ICLS resolution I) includes a more detailed set of status in employment categories defined in the International Classification of Status in Employment (ICSE-18) as compared with the previous standards. One of the major criticisms against the previous standards was that ICSE-93 did not provide sufficient information to adequately monitor the changes in employment arrangements taking place in many countries. The increase in non-standard forms of employment such as zero-hour contracts, short term contracts and "dependent self-employed" was not clearly addressed in ICSE-93 and challenged the boundary between self-employment and paid employment. The more detailed categories in ICSE-18 allow the provision of more comprehensive and internationally comparable statistics on the changing nature of employment relationships including the growth of non-standard forms of employment. It also provides a clearer boundary between independent workers and dependent workers as well as between workers in employment for profit and workers in employment for pay. The 20th ICLS resolution I also includes the Classification of status at Work (ICSaW-18) which is an extension of the classification of status in employment that covers all forms of work, paid as well as unpaid. The 20th ICLS resolution I is hence aligned to the 19th ICLS Resolution I concerning statistics of work, employment and labour underutilization, adopted in 2013 (ILO, 2013).

The classification is complemented by 23 cross-cutting variables that provide information for characteristics that cut across several status categories. Many of these are already regularly covered in Labour Force Surveys (LFS) but are not in general covered by international standards. The inclusion and definitions of these variables will thus promote greater harmonisation and international comparability. The resolution also contains a set of indicators that allow the assessment and monitoring of labour market performance, the stability of employment relationships, exposure of the employed population to economic risk and participation in non-standard employment arrangements. The 20th ICLS resolution I is a comprehensive set of standards where some of its aspects are familiar and already part of most countries LFS´s while others are new and will constitute a challenge to measure. The status in employment is still the core of the new standards and many countries are using the LFS as the main survey instrument for collecting status in employment categories. The focus in this paper is therefore on the classification principles and structure of ICSE-18 and how statistics according to the ICSE-18 can be collected in a typical LFS. The paper will cover the following main topics (2) ICSE-18, (3) Measurement of ICSE-18, (4) Conclusions.

2. International Classification of Status in Employment (ICSE-18)

ICSE-18 is at the core of the 20th ICLS resolution I. It comprises ten mutually exclusively statuses that refer to work relationships within employment. The classification uses two different aspects of the work relationship as criteria to differentiate categories of jobs according to their status: **the type of authority** that the worker is able to exercise in relation to the work performed and **the type of economic risk** to which the worker is exposed. The ten categories include four sub-categories of employees which allow identification of those with non-standard employment arrangements, four separate categories of independent workers and a separate category for respectively dependent contractors and contributing family workers. The category "workers in producers' cooperatives" previously included in ICSE-93 has not been retained in ICSE-18.

Dependent contractors is a new category in ICSE-18 and it addresses the challenge many statistical agencies have of how to define workers that are on the boundary between being employees and self-employed. The 20th ICLS resolution I defines dependent contractors as workers that have a commercial agreement to provide services or goods for or on behalf of another economic entity but at the same time are dependent (operationally or economically) on that entity for the organization and execution of the work or for access to the market.

One of the weakness with ICSE-93 was the blurry boundary between paid employment jobs and self-employment. Owner-mangers of incorporated enterprises and contractors were two categories of worker identified as "particular groups" in ICSE-93, these workers cut across two or more of the substantive groups and countries have to choose where to place them. This is addressed in ICSE-18 by the new structure of ten detailed mutually exclusive categories that can be organised not only according to type of authority but also according to type of economic risk. This creates two different hierarchies: the International Classification of Status in Employment according to type of authority (ICSE-18-A) and the Classification of Status in Employment according to type of Economic risk (ICSE-18-R).

ICSE-18-A reflects the degree of authority that the worker is able to exercise in the job. This creates a dichotomy between *independent workers and dependent workers*. Independent workers have control over how the work should be organised, can make the most important decisions about the activities of the business and are not economically dependent i.e. they do control access to the market, raw material and capital items. Conversely, dependent workers do not have complete authority or control over the economic unit for which they work and are accountable to or supervised by another person or economic unit. Two groups are classified under *independent workers,* namely, *employers* and *independent workers without employees*. The category *dependent workers* contains three groups, namely, *dependent contractors, employees* and *contributing family workers*. The distinctions made in ICSE-93 between employers, own-account workers, contributing family workers and employees are thus retained in this hierarchy. ICSE-18-A is suitable for various types of labour market analysis such as the impact of economic cycles, government policies and for the compilation of statistics classified by socio-economic status.

ICSE-18-R reflects to some degree the economic risk to which the worker is subjected in the job. This classification differentiates between workers in employment for profit and those in employment for pay. *Workers in employment for profit* are workers whose remuneration is dependent on the profit or loss of the economic unit in which they are employed. They thereby may face a greater economic risk than *workers in employment* for pay who receive remuneration in return for time worked or for each piece or service produced. Three groups are classified under *workers in employment for profit*, namely, *independent workers in household market enterprises, dependent contractors and contributing family workers*. The category *workers in employment for pay contains two groups, namely, owner operators for corporations and employees.*

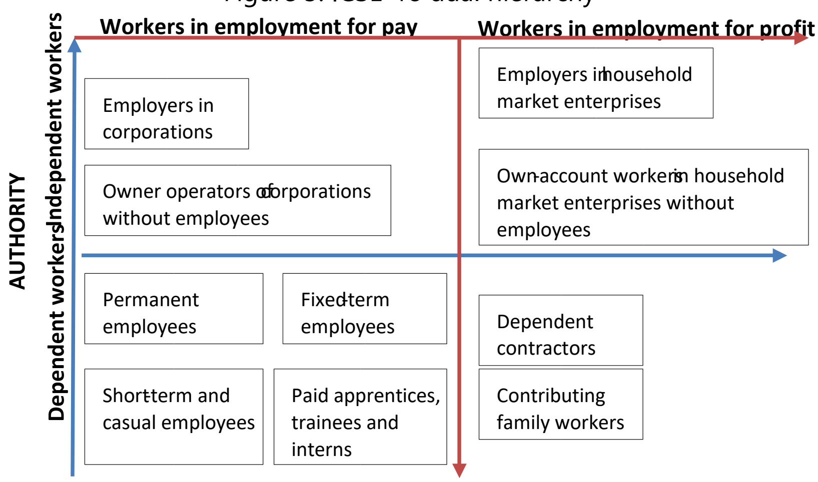
Figure 1: ICSE-18-A and ICSE-18-R ICSE-18-A

Independent workers
Employers:
-Employers in corporations
-Employers in household market enterprises
Independent workers without employees:
-Owner-operators of corporations without employees
-Own-account workers in household market enterprises without employees
Dependent workers
Dependent contractors:
-Dependent contractors
Employees:
-Permanent employees
-Fixed-term employees
-Short-term and casual employees
-Paid apprentices, trainees and interns
Contributing family workers:
-Contributing family workers

ICSE-18-R
Workers in employment for profit
Independent workers in household market enterprises:
-Employers in household market enterprises
-Own-account workers in household market enterprises without employees
Dependent contractors:
-Dependent contractors
Contributing family workers:
-Contributing family workers
Workers in employment for pay
Owner operators of corporations:
-Employers in corporations
- Owner-operators of corporations without employees
Employees:
-Permanent employees
-Fixed-term employees
-Short-term and casual employees
-Paid apprentices, trainees and interns

The possibility to organise the same ten different categories in two different hierarchies recognises that the relationship between the level of authority and level of economic risk is not always a direct positive relationship where a high degree of economic risk also implies a high degree of authority or vice versa. This might be the case for some categories. For example, employees have a relatively low degree of authority (as dependent workers) while experiencing a relatively low degree of economic risk (as workers in employment for pay). Moreover, own-account workers in household market enterprises have a higher degree of authority (as independent workers) and also a higher degree of economic risk (as workers in employment for profit). However, for other categories of workers there is an inverse relationship between authority and risk. For example, owner-operators of corporations without employees have a lower economic risk but a higher degree of authority while dependent contractors have a higher economic risk but a lower degree of authority. In practice, workers within these broad categories (as well as within the detailed categories) experience greater or lesser degrees of authority and economic risk and each of the two dimensions is rather a continuum than a dichotomy. For conceptual purposes, combining the two hierarchies (see figure 1) can however, contribute to a general understanding of the varying levels of authority and risk associated with the work relationship categories of ICSE-18 as well as a comparative framework to understand these dimensions for different types of workers.

Figure 3: ICSE-18 dual hierarchy



3. Measurement of ICSE-18

The 20th ICLS resolution I provides some general recommendations for data collection regarding data sources and frequencies of data collection. These general recommendations are not sufficiently detailed to give countries guidance on how ICSE-18 or other topics contained in the resolution in practice should be measured. To meet this need a more detailed set of guidelines (ILO, 2018b) was developed in parallel with the development of the 20th ICLS resolution. The data collection guidelines for ICSE-18 address the data collection of ICSE-18 and the cross-cutting variables when using household surveys (with particular focus on LFS), establishment surveys and administrative sources. The guidelines are based on current country practices and testing that has been conducted in relation to ICSE-18.

The proposed approach for collecting statistics according to ICSE-18 in a LFS is based on the structure in ICSE-18-A. The reason for this is that the aggregated second level categories in this hierarchy; *Employers, independent workers without employees, employees, dependent contractors and contributing family workers* are close to the five substantive categories in ICSE-93. To build the collection on the structure in ICSE-18-A thus allows countries to build on the approach already established for measuring ICSE-93. Most countries are currently collecting ICSE-93 by using one or two questions asking the respondent directly about their status in employment. This practice will not be sufficient to classify the status in employment as defined by ICSE-18. It will not allow identification of the detailed categories of ICSE-18 and it will not enable the identification of dependent contractors. The proposed approach seeks instead to use the existing practice as a starting point but to further build on it by identifying the necessary boundaries that are required for defining the detailed ICSE-18 category. The first step should as a minimum, capture self-identified self-employed, employees and contributing family workers. Depending on the outcome of the step, the respondent will be filtered to a classification module that defines the status in employment. In addition, a third module is necessary for those defined as employees to identify the four detailed categories of employees as defined in ICSE-18. The classification modules are captured in Figure 2 below. This modular structure ensures a flexible approach that can be used in different statistical sources with different levels of detail and frequency depending on the national context and need.

The different key-characteristics within the different modules can be turned into one question, or in a few cases two or more questions that can be directly integrated into an already existing survey. Many of the key-characteristics (such as having employees or not, forms of remuneration, characteristics related to the employment contract/agreement) are already part of the LFS in many countries. Nonetheless, some elements will be new for many countries.

Figure 2

Classification modules depending on self-identified (SI) employment status

SI- Self-employed

-Having employees or not
-Incorporated enterprise
-Operational and/or economic dependency

SI- Employees

-Forms of remuneration
-Responsibility for contribution to social insurance or deduction of income tax or actual contribution

SI-CFW

-Forms of remuneration
-Making most important decisions

Sub-classification module defined as Employees

Employees
-Guaranteed minimum amount of hours
-Time-limited employment arrangement
- Reasons for limited duration of employment agreement
- The duration of the employment agreement

Countries still have limited experience with measuring the new category of dependent contractors. The proposed measurement approach for identifying dependent contractors is based on two different tracks. The first track identifies dependent contractors among those persons that perceive themselves as being self-employed while the second starts by targeting persons that self-identify as employees. The two tracks are a recognition that within dependent contractors there are two sub-categories: workers who own and operate their own business but do not have full control or authority over their work, and workers who provide their labour to others but have contractual arrangements that are akin to self-employment. The first group would have a greater tendency to identify themselves as self-employed while the latter would more frequently identify themselves as being employees. It is hence necessary to identify dependent contractors among self-identified self-employed as well as self-identified employees in order to identify all dependent contractors.

The first track requires the operationalization of a boundary between own-account workers in household market enterprises and dependent contractors.

Employers and owner-operators of incorporated enterprises are by definition not dependent contractors. The boundary therefore needs to be established for those that have identified themselves as being self-employed and have responded that they do not have an incorporated enterprise nor do they have any employees. Different operational criteria such as having a dominant client that exercises operational control, dependency on another economic entity that sets the price for the products or services produced, and sharing profits with a third party have been tested and the results published (ILO, 2018d). The different tests have contributed to important insights regarding the first track approach, but more work is needed before a clear recommendation can be made regarding how this boundary should be operationalised.

The second track requires the operationalization of a boundary between employees that are not paid for time worked and dependent contractors. The conceptual boundary that distinguishes the two groups is whether there is a commercial agreement or an agreement of employment. The proposed operationalization of this boundary uses criteria concerning whether the perceived employer is responsible for withholding income tax or for arranging social insurance or not. If this is a responsibility of the employer (independently of whether the employer de facto fulfils the responsibility or not) then this indicates that there is an employer–employee relationship and the worker is defined as an employee. However, if this is a responsibility of a respondent who has initially self-identified as an employee, then this indicates a commercial relationship and the respondent is defined as a dependent contractor. The second track has been tested to a lesser extent but according to the limited findings, the approach seems promising (ILO, 2018d). The operationalization might nonetheless be problematic in countries with a high share of informality. In such countries, the proposal is to operationalise the boundary on the basis of actual contributions of social insurance in combination with other information such as place of work.

4. Conclusion

The labour force survey has been an important data source for collecting statistics by status in employment and will continue to be well suited to collect this information as now defined in ICSE-18. The proposed measurement approach is built around a self-identification module, which for most countries is the current approach to identify the ICSE-93 status. It would therefore be relatively easy to integrate the approach in most LFS. The different modules in a LFS would typically target the main job but could also be used in relation to secondary jobs. The modular approach creates a flexible structure that allows countries to adapt it to the structure of their specific survey as well as open up the possibility to measure the different levels of ICSE-18 with different frequencies depending on national needs and context.

The implementation of ICSE-18 will increase the number of questions in most LFS and hence raise respondent burden. However, it is important to keep in mind that many of the key characteristics are already often collected. In addition, most of the information needed for defining the ICSE-18 categories is also often quite relevant labour market information by itself.

The 20th ICLS resolution I is a clearer and more comprehensive framework that meets the needs of different types of users and that better reflects the economic risk and degree of authority that is associated with the work relationship. The expectation is that ICSE-18 will greatly improve the current situation where the work relationship is typically defined *solely* on the basis of the self-identification of the respondent and where the treatment of particular groups of workers is unclear, with the result that countries are using different practices. Instead, the new ICSE-18 requires a more objective measurement in which the status in employment is derived on the basis of information collected using multiple questions. This will contribute to increased international harmonisation. At the same time, the 20th ICLS resolution I also poses challenges for countries and more guidance is still needed. The proposed measurement approach gives clear guidelines that countries can use for most of the part of ICSE-18 and the cross-cutting variables. However, there are still some important aspects, such as the identification of dependent contractors that need further methodological testing and development of guidance. This paper has focused on ICSE-18 but there are also other aspects of the resolution that need further conceptual and methodological work.

The 20th ICLS resolution I and the data collection guidelines are important steps to improving the measurement of work relationships, but more work is needed and in the end it is a dynamic process that relies on the active participation from a broad and diverse set of countries. Countries are therefore encouraged to conduct further testing for contributing to the methodological work as well as carry out tests at the national level before implementing ICSE-18.

References

1. International Labour Organization, 2018a, *Resolution concerning statistics on work relationships*, adopted by the 20th International Conference of Labour Statisticians, Geneva. https://www.ilo.org/global/statistics-anddatabases/meetings-and-events/international-conference-of-labour-statisticians/20/WCMS_648693/lang-en/index.htm
2. -. 2018b, *Data collection guidelines for ICSE-18*, Geneva https://www.ilo.org/stat/Publications/WCMS_636039/lang--en/index.htm

3. -. 2018d, *Testing of proposals for a revised International Classification of Status in Employment*, Geneva, https://www.ilo.org/stat/Publications/WCMS_636046/lang--en/index.htm
4. -.2013. 19th *ICLS* r*esolution concerning statistics of work, employment and labour underutilization*, Geneva, https://www.ilo.org/global/statistics-and-databases/meetings-and-events/international-conference-of-labourstatisticians/19/WCMS_230304/lang--en/index.htm
5. -. 1993. *15th ICLS resolution concerning the International Classification of Status in Employment (ICSE-93)*. Geneva, http://www.ilo.ch/global/statistics-and-databases/standards-and-guidelines/resolutions-adopted-byinternational-conferences-of-labour-statisticians/WCMS_087562/lang--en/index.htm

Testing dependent contractors in the Danish Labour Force Survey

Tine Cordes, Thomas Thorsen
Statistics Denmark, Copenhagen, Denmark

Abstract

The new International Classification of Status in Employment (ICSE) from 2018 introduces a different view on the traditional dichotomy between self-employed and employees. Instead employment is divided both according to authority (dependent and independent workers) and according to risk (employment for profit and employment for pay). The dividing lines between these categories are not always clear-cut, which presents challenges for national statistical institutes when conducting their Labour Force Surveys (LFS). In particular, the category dependent contractors consisting of dependent workers who work for profit is presumed to be difficult to measure.

In the second quarter of 2017 the International Labour Organization (ILO) had some questions added to the Danish LFS in order to test their quality and their suitability to measure the concept of dependent contractors. Three tentative definitions of dependent contractors were constructed by ILO from these questions.

The groups found were relatively disjunctive, and the characteristics that were expected to be defining for dependent contractors did not seem to exist together very often. Moreover none of the three groups seemed to be as exposed, as one might expect to be the case for dependent contractors.

Even if the phenomenon of dependent contractors was widely acknowledged, an empirically proven delineation of dependent contractors was not established when the ICSE was adopted. Instead some potential characteristics were included in the adopted ICSE.

Revisiting the Danish test in this light, this paper has its main focus on those of the tested questions that are considered relevant in measuring dependent contractors according to the new ICSE. These include if another economic unit

- controls the price for the goods or services provided,
- controls working arrangements such as working time, or
- serves as a dominant customer.

The suitability of the specific questions regarding the identification of dependent contractors is discussed by presenting additional relevant characteristics of the respondents. Some of the characteristics are found in the LFS itself, while others are found by linking to administrative microdata.

Finally, the results are used to discuss the concept and measurability of dependents contractors in the relevant economic setting. Besides evaluating

the suitability of the tested questions for capturing the group of dependent contractors, additional questions that could be relevant for measuring the phenomenon in a Danish – or similar – context are considered.

Keywords
ICSE-18; gig economy; precarious employment; self-employment

1. Introduction

1.1. Dependent contractors as a phenomenon

The concept of dependent contractors is of high complexity embracing different types of formally self-employed exposed to economic risks and costs that self-employment implies, but without some of the advantages and liberties that normally also accompany being self-employed. Typical examples of dependent contractors could be vehicle drivers who search for rides through an online mobile telephone application, or cleaning workers who are hired formally as subcontractors, but who are in reality under full operational instruction of the contracting company as if the workers were employees. Another example could be an IT consultant working for a former employer – now on a business contract as self-employed. The consultant now works permanently in the premises of the former employer with the same tasks as before. In any case the worker may likely not be entitled to social benefits, legal protection, insurance etc., but on the other hand presumably has the right to deduct expenses in his or her tax report.

The phenomenon of dependent contractors is reported to be increasing, or at least creating more attention, and hence a description of dependent contractors has been included in ILO's new International Classification of Status in Employment from 2018 (ICSE-18). This resolution mentions a number of different and often quite general characteristics that may identify dependent contractors, and it seems yet too immature to offer an actual operationalization of dependent contractors. This paper seeks to contribute to this process by analyzing various potential indicators of being a dependent contractor in an economic setting like the Danish one. First, essential characteristics of dependent contractors according to ICSE-18 are described. Then the methodology of the analysis is briefly reviewed. Next the results of the analysis are presented and finally these results are discussed in light of the process of measuring the concept of dependent contractors.

1.2. Dependent contractors according to ICSE-18

It is clear in ICSE-18 that a dependent contractor is working for profit (in the same way as a classic self-employed), does not operate an incorporated enterprise and does not have any employees. However, these characteristics are also valid for Own-account workers in household enterprises without employees (from here on just own-account workers). The difference is that the dependent contractor is operationally or economically dependent on another economic unit. The further explanations contained in the ICSE-18-definition of the concept are possible characteristics or conditions, not based solidly on either a theoretical basis or a proven empirical identification, but more loosely on observations of real-life phenomena, such as platform work, pro forma self-employment and home assembly. Such phenomena are of a very diverse nature, not only in different types of economies, but sometimes even within a single national economy.

2. Methodology

2.1. The 2017 test of dependent contractors

As part of the preparation for ICSE-18, ILO had some questions added to the Danish LFS in the second quarter of 2017. The questions were meant to try to capture main aspects of the description of a dependent contractor. The questions were integrated into the 2017 ad hoc module on self-employment developed by Eurostat. Indicators of economic and operational dependency were subsequently chosen from the module and grouped by ILO into three different tentative definitions of dependent contractors. All three definitions were delimited to self-employed without employees who did not own an incorporated company (the test group). According to the first definition a dependent contractor was characterized by the fact that at least 75 percent of income came from the same client and the client controlled the working time of the worker. According to the second definition a dependent contractor was characterized by not controlling price-setting of his or her products. According to the third definition a dependent contractor was characterized by sharing profit with a third party. In the Danish test definition three had a very large part of respondents which was not part of any of the other two definitions, and moreover the group did not seem to be exposed in any way. Furthermore, test results from other countries also suggested problems with this definition, and it was consequently dropped from being mentioned in the resolution. Therefore this definition will not be looked into in this paper. Instead the results of the remaining two definitions will be reviewed.

2.2. Further Analysis

The 2017 ad hoc module contained more questions related to limitations in economic or operational liberties for self-employed and to working arrangements that may be similar to those of employees. Such questions may indicate that the worker could be a dependent contractor. Among these potential indicators we include the questions used in the first two ILO-test definitions. The set of indicators of being a dependent contractor is then expanded to:

Indicator	Potential dependency
Only one client or at least 75 percent of income from the same client during the last 12 months	Operational or economic dependency of client
Clients determine working hours fully or in part	Operational dependency of client
Price is set by another company or agent or by the client	Economic dependency of client or other
Former employer encouraged the person to become self-employed	Operational or economic dependency of client
Not enough influence with regards to pricing work	Economic dependency of client or other
Not possible to have employees within field of work	Operational or economic dependency, working conditions similar to those of employees
Clients want the worker to carry out the work personally	Operational or economic dependency, working conditions similar to those of employees
Worker is able to influence the content of work-assignments to a lesser extent or not at all	Operational dependency, working conditions similar to those of employees
Worker is able to influence the order of work-assignments to a lesser extent or not at all	Operational dependency, working conditions similar to those of employees

Furthermore, we identify questions that may indicate that the worker is an own-account worker. The idea is that some of these may disqualify the worker from being a dependent contractor. These indicators are:

1. Typically sets the price charged for products or services
2. Plans to employ people within 12 months
3. Plans to use subcontractors within 12 months
4. Worker is able to influence content of work assignments to a great extent
5. Worker is able to influence order of work assignments to a great extent

To assess the strength of the individual indicators in identifying dependent contractors, we divide all self-employed in the test group into two subgroups: Workers who might be dependent contractors, and workers who are unlikely to be dependent contractors. This is done by using information from the core LFS, such as descriptions of company, activity, occupation, tasks and types of clients of the worker. This manual division of the workers is then cross-tabulated with the two sets of indicators.

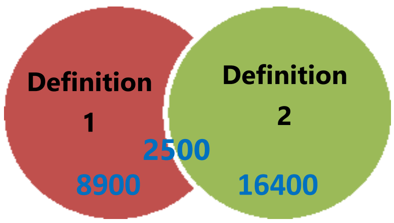
We also use information from administrative sources (Statistics Denmark's Business Register and Income Statistics) to cross-tabulate with the dependent contractor indicators in order to get more information on those workers who are captured by each of the indicators.

3. Result:

3.1. Test results from the Danish test carried out for ILO

In the test conducted for ILO, definition 1 – dependency of a client – counted 11,300 persons representing around 5 percent of all self-employed and 9 percent of the self-employed in the test group (no incorporated company, no employees). Compared to the remaining part of the test group, the group meeting definition 1 seemed to have a specific profile around consultants and highly educated professionals. They showed indications of being slightly more exposed due to lower turnover and working time, but the small size of this group made it impossible to draw any conclusions. Definition 2 – the price control approach – comprised 18,900 persons which was 8 percent of all self-employed and 15 percent of the self-employed in the test group. This group did not show signs of exposure, but rather captured a certain industry, namely the farming industry, where price setting often is outside control of the worker. Hence, inability to set prices seems inadequate by itself, but might rather be an indication of dependency when co-existing with other dependency-indicators, for example dependency of clients. However, as can be seen from Figure 1, dependency of clients and the price control approach did not seem to co-exist very often. Only 2,500 persons, corresponding to 9 percent of those meeting any of the two criteria, met both of the two criteria. The 2,500 persons are just 2 percent of all self-employed in the test group and hence too small a group to base any conclusions on.

Figure 1. Overlap between definitions of Dependent Contractors. Persons



As indicated above neither of the two definitions by themselves seemed to capture people in especially precarious, dependent situations. That might be because neither of the characteristics involved in the definitions seem to be unique to dependent contractors, as many ordinary businesses work under similar conditions without being particularly dependent, let alone in a particularly weak position. On the contrary, some of the limitations may indeed increase business opportunities:

- Limiting oneself to an agreement with one single client on a fixed deal may provide a more stable and sometimes higher income. This is probably true for some groups. If there are many potential clients and the market situation is favorable, having only one client at the time may not be particularly limiting.
- Adhering to a common price concept, branding agreements or similar may be attractive. This could include marketing advantages, so that the potential number of clients and even prices could be higher.
- As the two definitions do not seem to capture the target group – neither independently nor in combination, and as ICSE-18 contains a broader definition than the two tested, we extend the number of indicators as mentioned under methodology.

3.2. Revisiting the test

Table 1 shows the different number of self-employed persons in the test group, who have answered yes to the specific dependent contractor indicators, and cross-tabulated with likeliness or not to be a dependent contractor via a manual examination of the respondents in the whole test group.

In the manual division of the self-employed in the test group, 35,600 are deemed potential dependent contractors, while 88,200 are deemed unlikely to be dependent contractors. Of the 35,600, 10,900, or almost a third, are in NACE section M Professional, scientific and technical activities, while 5,300 are in section J Information and communication. Cross-tabulating this with the indicators, also in table 1, we see that 59 percent of workers having been encouraged to become self-employed by former employer was deemed potential dependent contractors according to manual check. This might be an indication of economic or operational dependency of a client, presuming the former employer could still be a client.

Other dependency indicators with relatively high shares of manually identified potential dependent contractors are: Single dominant client and Clients determine working time that in combination constituted the main client approach in the 2017 test carried out for ILO. Relatively high shares having these indicators are in this further test manually categorized as likely to be dependent contractors, and this give strength to the conclusion from 3.1. regarding that these conditions might be indicators of dependency.

Furthermore, Employees not possible in field of work and No/little influence on order of assignments could be indicators of dependency as almost half of respondents who answered yes to these questions were potential dependent contractors according to the manual division. Such circumstances indicate reduced influence and freedom compared to what is typical for a classic self-employed.

The indicators regarding pricing are quite weak, supporting the conclusion from the 2017 test. Those with the indicator price set by others are even less likely to be potential dependent contractors than those without this indicator. However, it is to be noted that some of the numbers are quite small and conclusions must be taken with care. Importantly, none of the indicators have an overwhelming share of potential dependent contractors. Thus, no single indicator is by itself enough to define a dependent contractor.

Table 1. Indicators of being a dependent contractor.
Number with indicator true/false, percentage deemed potential dependent contractors

	Number of persons		Percentage potential dependent contractors according to manual division	
	Indicator true	Indicator false	Indicator true	Indicator false
	Persons		Percentage	
1 Single dominant client	36 200	87 600	41	24
2 Clients determine working hours	32 700	91 100	41	24
3 Price set by others	18 900	104 900	17	31
4 Former employer encouraged self-employment	8 000	115 800	59	27
5 Not enough influence on pricing	25 400	98 400	30	28
6 Employees not possible in field of work	12 000	111 800	46	27
7 Clients want work carried out personally	52 000	71 800	33	25
8 No/little influence on content of assignments	4 400	119 400	.	30
9 No/little influence on order of assignments	6 500	117 300	44	28

Table 2 regards indicators that the worker is not a dependent contractor (and therefore an own-account worker). It shows that especially plans of hiring employees or subcontractors are often ruling out the possibility of being a dependent contractor. None of the other indicators seem to work in particular against being a dependent contractor.

Table 2. Indicators of being an own-account worker.

Number with indicator true/false, percentage deemed potential dependent contractors

	Number of persons		Percentage potential dependent contractors according to manual division	
	Indicator true	Indicator false	Indicator true	Indicator false
	Persons		Percentage	
Setting price,	89 300	34 600	29	29
Planning hiring employees	22 600	101 200	17	31
Planning subcontractors	28 900	94 900	16	33
High influence on content of assignments	99 300	24 500	31	19
High influence on order of assignments	92 000	31 800	28	30

Table 3 shows the median annual personal income for self-employed in the test group for each of the dependent contractor indicators. It shows that for several indicators, the median income and median turnover is higher for those answering yes to the indicator than for those answering no. This implies that the indicators do not necessarily catch workers in especially vulnerable positions within the test group. The test group as a whole, though, has relatively low income and turnover compared to other types of self-employed. It is to be expected that persons with higher degree of exposure will be a significant part both among dependent contractors and own-account workers.

Table 3. Indicators of being a dependent contractor. Median income and turnover

	Median personal income DKK		Median annual turnover DKK	
	Indicator true	Indicator false	Indicator true	Indicator false
	Persons			
1. Single dominant client	261 000	232 000	264 000	363 000
2. Clients determine working hours	283 000	234 000	272 000	358 000
3. Price set by others	219 000	255 000	573 000	264 000
4. Former employer encouraged self-employment	294 000	245 000	.	358 000
5. Not enough influence on pricing	190 000	255 000	413 000	287 000
6. Employees not possible in field of work	319 000	234 000	386 000	312 000
7. Clients want work carried out personally	303 000	224 000	207 000	388 000
8. No/little influence on content of assignments	115 000	255 000	195 000	353 000
9. No/little influence on order of assignments	276 000	245 000	195 000	353 000

Summing up on these different tests, it seems that no single indicator alone captures dependency. Thus, a number of indicators will have to be present to qualify as a dependent contractor, while other indicators may disqualify. Indicators regarding pricing seem unable to capture dependent contractors effectively, while the other indicators may have potential. It is likely that those who are identified as dependent contractors are not more exposed than others in the (already relatively exposed) test group.

4. Discussion and Conclusion

4.1. Conclusions with discussion

We may conclude the following regarding dependent contractors in an economic setting like Denmark:

1. Any single criterion used for identifying a dependent contractor, seem to include many workers who are not likely to be dependent contractors – probably more than half. Therefore workers should most likely meet more than one criterion to qualify as a dependent contractor. The composition of such criteria needs more work.
2. From the results it seems that using **the price control** approach is not a particularly viable way of identifying dependent contractors. In the 2017 test, the groups that were identified were deemed not sufficiently relevant, and in the new test, the manual check showed that the share of potential dependent contractors was very low. The failure of these criteria might be due to the fact that many self-employed in the test group (no incorporated company, no employees) – even if they do not depend on another economic unit – have little influence on the market price. Even if the self-employed could in theory set the price of the goods and services provided, this may not be possible in reality due to market factors.
3. Criteria about having **a single dominant client** may, according to the results, work in some cases, but like any other indicator, not by itself (cf. conclusion No. 1).
    a. The concept of client is unfortunately ambiguous in many cases regarding capturing potential dependent contractors. The presumed client may in some cases consider persons as his or her own clients when they are in fact the clients of the employing company. This could lead to misclassification.
4. It seems like a good idea to focus on **working arrangements and conditions that are similar to those of employees,** a point that is also mentioned in ICSE-18. Initially, it is arguably a good theoretical starting point that could also serve as an umbrella for other characteristics of dependent contractors. Moreover, the results show that indicators of working arrangements and conditions that are similar to those of employees have relatively high shares of potential

dependent contractors compared to other indicators according to our manual division.

a. Even if the contract between a dependent contractor and the employing company is on the face of it a business contract, the job is presumed to be carried out personally. The worker not being **able to hire an employee or a subcontractor to do the work** could be used when identifying a dependent contractor. Thus, the question [having] employees not possible in field of work had a relatively high share of potential dependent contractors, nearly half. Even planning to hire an employee or subcontractor could on the other hand disqualify from being a dependent contractor. Clients want work carried out personally had a share of potential dependent contractors higher than those who did not have the indicator. Furthermore the indicators of not being a dependent contractor, planning hiring employees and planning subcontractors had very low shares of potential dependent contractors.

b. **The employing company controlling the business processes** (organizing or supervising the work) and having the right – even if not exercising it – to define the content of the tasks, or prioritize the sequence of them, and to make important decisions about the operation may also be useful in capturing dependency. If the work of the self-employed is integrated into the business of the employing company rather than being peripheral to it, so the self-employed might be operationally dependent on the employing company. Almost half of those in the test group who had no/little influence on order of assignments were potential dependent contractors according to the manual check.

5. When looking at the income and turnover of potential dependent contractors, it is **not very likely that dependent contractors are particularly more vulnerable** than the rest of the test group (own-account workers), regardless of operationalization. The test group as a whole is relatively vulnerable compared to other self-employed, and the vulnerable workers are likely to be present among both dependent contractors and own-account workers.

4.2. Summing up and further work

As regards the process of operationalizing the concept of dependent contractors in an economic setting like the Danish one, it seems like a good starting point to divide the test group into potential dependent contractors and unlikely dependent contractors. Doing this enabled us to better assess the strength of each indicator. Cross-tabulation this division with the indicators showed that variables measuring working arrangements and conditions that are similar to those of an employee could be a good starting point for operationalizing. Further cross tabulations with other indicators would be wise to come closer to an actual operationalization, but this would probably entail asking more respondents than was done in the 2017 test, e.g. the full LFS sample.

Further work could be done on the theoretical framework, as to facilitate the identification and understanding of the concept of dependent contractors. Moreover, more tests of yet untested indicators entailed in ICSE-18 could usefully be carried out, such as whether the employing company is supplying work facilities, equipment or tools (which the dependent contractor sometimes must pay for using), and also whether the financial reward-risk balance is similar to that of an employer-employee relationship or even skewed in favour of the employing company. Finally, work could be done to shed light on how indicators could combine to establish an operationalization of the identification of dependent contractors.

References

1. Cordes, T. & Thorsen, T. (2017): Dependent Contractor Questions in LFS. Unpublished.
2. International Labour Organization (2018): Resolution concerning statistics on work relationships. www.ilo.org

Entrepreneurs, multiparty work relationships and work in the internet-based "Gig Economy": How to advance on statistical measurement?

Monica D. Castillo, Manpreet Singh
International Labour Organization, Department of Statistics, Geneva, Switzerland

Abstract
In October 2018, the 20th International Conference of Labour Statisticians (ICLS) adopted the Resolution I concerning statistics on work relationships (henceforth, "the Resolution") which builds on previous statistical standards related to status in employment, establishing a more refined and far-reaching framework intended ultimately to support evidence-based research and policy regarding the complex and changing world of work relationships. The Resolution will support the production of national statistics on work relationships and facilitate international comparability. It contains a set of priority topics that will require further conceptual and methodological development and also areas of work identified in the future work agenda. This paper focuses on three such topics, namely: multiparty work relationships, intermediated platform work (sometimes referred to as work in the Internet-based "gig economy") and entrepreneurs. Multi-party work relationships and intermediated platform work are often associated with non-standard forms of employment and are therefore of high policy relevance. Entrepreneurs represent a subcategory of independent workers and include both own-account workers and employers in both incorporated and unincorporated enterprises. The paper provides information on the policy interest from a decent work perspective for the topics as well as concept definitions. The paper also describes key objectives, challenges and approaches to support further conceptual and methodological development in order to help countries in measuring these topics.

Keywords
Multi-party work relationships; Entrepreneurs; Intermediated platform work; Internet-based gig economy; International Conference of Labour Statisticians

1. Introduction

In October 2018, the 20th The International Conference of Labour Statisticians (ICLS) adopted the *Resolution I concerning statistics on work relationships* (henceforth, "the Resolution") which built on previous statistical standards related to status in employment. The Resolution established a more refined and extensive framework intended to support evidence-based research and policy regarding the complex and changing world of work relationships. It is intended to support the production of national statistics on work relationships and facilitate international comparability. A suite of standards was established under the Resolution that included a conceptual framework; a key classification on status in employment (revised from 1993 and referred to as ICSE-18) and on status at work; operational concepts, definitions and guidelines; as well as a set of cross-cutting variables and categories.

The Resolution contains a set of priority topics that will require further conceptual and methodological development and also areas of work identified in the future work agenda. This paper focuses on three such topics, namely, multi-party work relationships (MPWR), intermediated platform work (IPW, sometimes referred to as work in the Internet-based "gig economy") and entrepreneurs.

The three selected topics represent areas of policy interest in the context of promoting decent work. MPWR and IPW share some common features. Both can be considered distinct types of third-party relationships that offer the possibility of employment to a variety of workers across the globe in different industries and occupations. Many workers that choose IPW or temporary agency work (a subcategory of MPWR) may benefit from the flexibility of working time arrangements given other demands in their lives (e.g. school studies or family care).

At the same time, MPWR and IPW are associated with the concept of "non-standard employment" (NSE). According to a 2016 ILO report[1], NSE comprises employment arrangements that deviate from the "standard employment relationship", understood as work that is full time, indefinite, as well as part of a subordinate relationship between an employee and an employer. The 2016 report presents information suggesting that subcategories of multi-party employment relationships are often characterized by various types of insecurities and risks, including extended working hours, occupational safety and health risks, training insecurity, and representation insecurity as well as labour market transitions and employment insecurity, penalties in earnings, and inadequate social security coverage[2]. The 2018 ILO report on "Digital labour platforms and the future of work"[3] similarly highlights the insecurities experienced by microtask platform workers, noting that such workers experience low earnings, flexible work schedules with atypical hours, skill mismatch and lack of career advancement, and lack social protection benefits. Some of these workers express frustration over one-sided rating systems, an

---

[1] ILO, 2016. Non-standard employment around the world: Understanding challenges, shaping prospects, Geneva, p. xxi.

[2] Ibid. See pp. 185-228; figure 5.11 p. 224.

[3] ILO, 2018. Digital labour platforms and the future of work: Towards decent work in the online world, Geneva

inability to communicate with requesters and platforms and work rejected and payment refused.

Regarding entrepreneurs, the decent work agenda provides a valuable framework for considering effective ways for supporting entrepreneurs who create employment for themselves and potentially for others. Entrepreneurship can be a channel for instituting mechanisms that provide greater protection and security to entrepreneurs and their families and communities. Within a targeted decent work policy framework, entrepreneurs (including marginalized women and youth entrepreneurs) can become aware of their rights and entitlements to available resources. This can contribute to combating discrimination in the labour market.

This paper seeks to present information to support further conceptual and methodological development on the topics of MPWR, IPW and entrepreneurs that can eventually lead to guidance on statistical measurement. The remainder of the paper is structured around the following sections: concept definitions; objectives, challenges and approaches; and conclusion and discussion.

2. Concept definitions

Further development of the conceptual and methodological framework on the three selected topics must start with the adopted concept definitions or, in the absence of such a definition, a general understanding of the issue. This section provides an overview of definitions for each topic covered in the paper.

Multi-party work relationships. The Resolution defines **multi-party work relationships** as those involving a third party between a dependent worker and the enterprise for which the work is performed (the user enterprise). Workers in this category are by definition dependent and thus include employees and dependent contractors. It should be noted, however, that while contributing family workers (CFW) are also a subcategory of dependent workers, they are not part of MPWR since by definition CFW must assist a family or household member in their job (either in a market-oriented business operated by the other member or in another job as an employee or dependent contractor) and they are paid via intra-household transfers. In the case of **dependent contractors with MPWR**, a third party (that is, the economic unit with which the dependent contractor has a commercial contract to produce the goods or services but which does not receive the output produced by the worker) controls the worker's access to raw materials, clients or the market. Dependent contractors may be paid directly by the client, or payment may be received through an intermediary that benefits from the work performed.

In the case of **employees with MPWR**, the third-party enterprise serves as the employer, responsible for hiring and paying the wages to the workers while also mediating the temporary or permanent relationship between the employee and the user enterprise. The user enterprise is responsible for supervising the work of the employees. Moreover, the main place of work is usually the premises of the user enterprise.

There are two main groups of employees with MPWR, namely **agency workers** and employees providing outsourced services. In some countries, a third group may also be identified, namely workers in employment promotion schemes. According to the Resolution, agency workers are defined as dependent workers supplied by an agency to work for another enterprise under the supervision of the user enterprise. They are employed by private employment agencies, such as labour hire agencies, temporary employment agencies, or others that hire and make workers available to user enterprises but do not supervise the work. **Employees providing outsourced services** are hired as employees by one enterprise to provide, on a regular basis, specific services that their employer has contracted to provide to another enterprise or to a household, at the premises of and usually under the partial supervision of the client. Employers may include for example nursing agencies, domestic or office cleaning service providers, security service providers and information technology services providers. **Workers in employment promotion schemes** are workers provided and paid by a government agency to perform work for another economic unit as part of a government-funded employment promotion programme.

Intermediated Platform work. While there is no recognized international statistical standard definition of **intermediated platform work (IPW)** or similar concepts, it is thought in general to refer to employment that is organized or mediated through an Internet platform when the worker is not an employee (from a statistical measurement standpoint) of the enterprise operating the platform.[4] Other terms that have been used to designate this type of work in recent reports or by the media are *digital labour platform work, gig economy work and electronically mediated work.*

At this early stage of the research, it seems that the status in employment categories associated with IPW could be dependent contractors and employers, but not employees or contributing family workers. However, further research is necessary to fully substantiate this. The outcome will have direct implications on the measurement approach in surveys that identify the status in employment.

[4] ILO, 2018. Conceptual framework for statistics on work relationships, Room Document 5 presented to the 20th International Conference of Labour Statisticians, Geneva.

IPW is carried out in a triangular relationship in which a worker agrees to complete a work output or task via the digital platform in exchange for payment. The task request originates with a so-called "requester" or client that uses the platform to engage a worker (or workers) to complete the work. The nature of tasks or work output differs widely as do the skills required to perform them, and digital labour platforms function as microtask, macrotask or software development platforms. The payment structure for workers is based on successful completion of the task or work output and not on the time worked. Algorithmic management, which controls many aspects of the work performed, is a defining characteristic of IPW and involves work assigned, optimized and evaluated through algorithms and tracked data.[5] Task requesters or clients often anonymously rate the work of individual workers via the platform, with important consequences for the worker's ability to secure future task work on the platform.

IPW is heterogeneous and includes two main types: web-based work and location- and app-based work.[6] **In web-based work**, specific tasks are either outsourced to a crowd (i.e., a large group of global workers who often span different geographic and time zones) or directly to individual workers using a freelance marketplace. In **location- and app-based work**, tasks are primarily assigned to individual workers, for example in transportation, delivery and household services with fewer opportunities for the crowd.

Entrepreneurs. Economic and labour market policy analysts use statistics on status in employment to identify entrepreneurs and to assess the impact of self-employment and entrepreneurialism on employment and economic growth. According to the Resolution, **entrepreneurs** are persons who own and control an enterprise and seek to generate value through the creation of economic activity, by identifying and exploiting new products, processes or markets. In doing so, they create employment for themselves and potentially for others. Entrepreneurs represent a subcategory of independent workers and include both own-account workers and employers in both incorporated and unincorporated enterprises. Additional information relevant to the national context, such as the size, age and other characteristics of the enterprise, is needed to provide complete statistics on entrepreneurship and to accurately identify those workers who are creating employment opportunities for themselves or for others. **Entrepreneurship** in this paper refers to the phenomena associated with entrepreneurial activity.

[5] ILO 2018. Digital labour platforms and the future of work: Towards decent work in the online world, Geneva, p. 9.
[6] Ibid, p. 4.

3. Objectives, challenges and approaches

This section presents the key objectives, challenges and approaches associated with each topic discussed in this paper to help support further conceptual and methodological development.

Multi-party work relationships. The objective going forward should be to develop more solid measurement approaches and carry out testing for all three types of employees in MPWR (i.e., agency workers, employees providing outsourced services and workers in employment promotion schemes) in order to support development of data collection guidelines on measuring MPWR that would include model questions. This effort should initially target labour force surveys and subsequently other complementary data collection instruments. The proposed measurement approaches contained in the room document prepared prior to the 20th ICLS should be reviewed to evaluate to what extent they remain valid considering not only the modified definition of MPWR adopted during the conference but also the need to separately identify employees and dependent contractors among the dependent workers in MPWR. Agency worker employees are already readily measured in different country contexts which could allow for developing concrete testing proposals in labour force surveys relatively quickly (with the support of willing countries). Developing guidance on the measurement of employees providing outsourced services and workers in employment promotion schemes will require further review and analysis of national practices in order to develop sound approaches and start testing in countries.

Moreover, developing approaches for measuring dependent contractors in MPWR will need to be coordinated and integrated with the methods development and testing of the dependent contractor group as a whole. Following the development of initial data collection guidelines targeting labour force surveys, a review of country practices used to measure such workers using additional data sources including establishment surveys and administrative records could be very useful to support countries to measure and monitor decent work among workers in MPWR using different sources. Future work should also seek to develop guidelines to help measure third-party hiring enterprises and client enterprises engaged with workers in MPWR, to help understand their incidence and characteristics.

Intermediated Platform work. The immediate objective is to define the scope for statistical measurement as well as review and document existing practices, measurement methods and tools at the national and international level. This will involve conducting a thorough review at the global level (including both developed and developing countries) to better understand the types of IPW that exist across countries (including web-based work and location- and app-based work; and microtask, macrotask and software development platform workers), concepts and definitions used, and existing methods of statistical measurement. The initial focus would be on measuring worker incidence in IPW as well as working conditions especially through a labour force survey or targeted household survey, for example, focusing on households with Internet access. As the issue of connectivity is central to the concept of IPW, this must be taken into consideration when developing the sampling frame particularly in countries (and subregions such as rural areas) where Internet access may be scarce.

Work will also focus on reviewing additional data collection approaches such as specialized surveys targeting digital labour platforms to obtain information regarding workers in IPW or targeting Internet users. The benefit of conducting surveys directly with the digital labour platforms is that operators are able to provide information regarding the parties in the triangular relationship, including the number of registered and active platform workers; types of tasks; skills requirements, working time aspects of the tasks, payment to workers by the task; functioning of algorithmic management to assign, monitor and evaluate completion of tasks and data tracking; number and type of requesters or clients and volume of client requests over a given reference period. Special surveys could seek to obtain information about the number and characteristics of platform operators and the requesters that use the platform. However, it’s important to bear in mind that IPW is a global phenomenon in which workers, clients, and platform operators are often geographically dispersed across countries and even continents, posing unique challenges for measurement, suggesting the need for cross-border cooperation on data collection.

The findings will help support development of a proposal for a statistical conceptual framework and preliminary measurement approaches using labour force surveys and other household surveys as well as other specialized data collection instruments and methods that could be tested. Such a proposal could be crafted to support the establishment of a cross-cutting variable on IPW in the context of measuring work relationships, supporting the Resolution. Alternatively, it could be used to develop a stand-alone conceptual framework and methods for measuring IPW that would be developed as a proposal for a new statistical standard on IPW, being fully aligned with ICLS Resolutions on work statistics and work relationships.

Entrepreneurs. The approach to measuring entrepreneurial activity has evolved significantly over time. The Resolution addresses the limitations of the System of National Accounts (SNA)-based definition for the identification of entrepreneurs.

The ICSE-18 moves away from the term self-employed previously used in ICLS standards since it is interpreted differently in SNA as compared with labour market statistics. ICSE-18 introduces the category of independent workers, who are those employed persons who own and control an economic unit for which they work, whether it is incorporated or not. The category of independent workers in the classification of status in employment would provide the best starting point for the identification and compilation of statistics on entrepreneurs.

Comprehensive and internationally comparable data for entrepreneurial activity is lacking. One of the main challenges is that data on entrepreneurship span a multitude of domains. They cross the traditional boundary between economic and social statistics by encompassing topics such as the demography and performance of businesses, the profiles of individuals who create enterprises, the attitudes toward entrepreneurship, and the regulatory environment for setting up new businesses.

The complex nature of entrepreneurship and its significance for economic growth demand internationally comparable indicators that will enable distinguishing entrepreneurship from other business activities. The aim moving forward is to develop practical guidance on the identification, measurement and compilation of statistics on entrepreneurs. The level of detail may vary depending on the statistical source and on descriptive and analytical needs. Different statistical sources (household-based surveys, establishment-based surveys, and administrative sources), have distinct advantages as well as limitations and can frequently be complementary to each other. A review of current country practices and existing methods of statistical measurement is expected to be undertaken.

4. Conclusion and discussion

The topics of MPWR, IPW and entrepreneurs are relevant for the future of work globally and require further conceptual and methodological development work with a view to supporting countries with statistical measurement. The topics of MPWR and IPW share common elements but are also quite unique and all three topics will require targeted approaches to measurement. Of the three topics discussed in this paper, IPW is the only one for which an international concept definition hasn't yet been established, thus moving forward on defining the scope of a proposed definition will be critical. In all cases, consultations with stakeholders will be essential during the development process to ensure alignment of information needs of policymakers and researchers on the one hand, and soundness of proposals concerning statistical methods from the perspective of data producers, on the other. Consultations will ensure adequate support for the proposals, thus allowing proposed approaches and testing to go forward particularly where the concept definition is established in the international standards, i.e. for MPWR and entrepreneurs. Suggestions and refinements to the proposed set of objectives, challenges and approaches in support of further conceptual and methodological development presented for each topic in this paper are most welcome.

References

1. International Labour Organization, 2016. Non-standard employment around the world: Understanding challenges, shaping prospects, Geneva. https://www.ilo.org/global/publications/books/WCMS_534326/lang--en/index.htm
2. International Labour Organization, 2018. Resolution concerning statistics on work relationships, adopted by the 20th International Conference of Labour Statisticians, Geneva. https://www.ilo.org/global/statistics-and-databases/meetings-and-events/international-conference-of-labour-statisticians/20/WCMS_648693/lang--en/index.htm
3. International Labour Organization, 2018. Conceptual framework for statistics on work relationships, Room Document 5 presented to the 20th International Conference of Labour Statisticians, Geneva. https://www.ilo.org/stat/Publications/WCMS_636041/lang--en/index.htm
4. International Labour Organization, 2018. Multi-party work relationships; concepts, definitions and statistics, Room Document 9 presented to the 20th International Conference of Labour Statisticians, Geneva. https://www.ilo.org/stat/Publications/WCMS_636045/lang--en/index.htm
5. International Labour Organization, 2018. Digital labour platforms and the future of work: Towards decent work in the online world, Geneva. https://www.ilo.org/global/publications/books/WCMS 645337/lang--en/index.htm

Testing of dependent contractor sub-category in the Thailand Labor Force Survey 2017

Wissanee Poonsab[1], Françoise Carré[2], Joann Vanek[2]
[1]Thailand National Statistical Office
[2]WIEGO

Abstract

Using the Thailand 2017 Labor Force Survey, a test was carried out of questions designed to identify Dependent Contractors, a sub-category of Dependent Contractors recommended in the 20th ICLS Resolution on Work and Employment Relationships. The primary focus of the test was the identification of dependent contractors among self-identified self-employed workers. Regarding identifying likely dependent contractors through this *self-employed path*, we find that combining the "price control" criterion with the criterion of getting supplies "from others" provides a good, if narrowly defined, indicator of dependent contractor status for unincorporated own-account self-employed workers, especially those outside of Agriculture. By itself, the price control criterion also is a fairly good indicator. Important next steps would be to examine separately the Agriculture and non-Agriculture workers who do not control price and "get their own supplies." Conclusions provide suggestions for additional questions.

Keywords

Economic dependency; Operational dependency

1. Introduction

The analysis presented in this paper is the result of a collaboration between the Thailand National Statistical Office and WIEGO, in consultation with the International Labour Office Department of Statistics. WIEGO is a global research and policy network focused on improving the livelihood of workers in the informal economy, especially women. WIEGO places priority on the development of statistics on these workers as a crucial element in polices and plans to improve their livelihood. At the request of HomeNet South Asia the Thailand Statistical Office and WIEGO collaborated to prepare statistics on home-based workers –both independent home-based workers and dependent homeworkers - through data in the national labour force survey. To do so, questions were added to the 2017 Labour Force Survey. These questions also provided a test for the criteria required to identify dependent contractors.

The primary focus of the test was the identification of dependent contractors among self-identified self-employed persons. Often home-based workers are identified as self-employed when given the arrangements in which they work, they should be classified as dependent homeworkers. Dependent contractors are a sub-category of Dependent Workers that has been recommended by the 20th ICLS Resolution on Statistics on Work and Employment Relationships (ICLS 2018).

2. Methodology

The questions included targeted self-identified self-employed persons without employees[1], allowing the assessment of the following two different boundaries for identifying dependent contractors:

(a) Whether the worker receives raw materials or instructions provided by "customer", or another company, agent, or contractor. (indicates operational dependency)

(b) Whether the pricing of their work/ product is set by a third actor or customer. (indicates economic dependency).

A condition also applied selected only the own-account self-employed whose enterprise is unincorporated. However only a small number of own-account self-employed consist of an incorporated enterprise.

The data are from the Informal Employment Survey of the Thailand Labour Force Survey conducted yearly on a quarterly basis. This module was conducted throughout the country in July, August and September 2017. The data collection was face-to-face interview with the head of household or members of the household. The reference week is 7 days before the interview date.

3. Results

The analysis indicates there is an overlap between <u>not</u> controlling price and getting supplies or instructions from others. Fourteen per cent of workers who do not control price also receive supplies or instructions from "customer(s)" or "another company/agent/contractor" (Table 1). Therefore, in the aggregate, the variable of "price control" by itself points to a large share of workers most likely to be dependent contractors

[1] The other sub-categories of self-employed worker—Employers and Contributing Family Workers—were not asked these questions; it would provide useful information

Table 1. Criteria for identifying Dependent Contractors among the self-identified self-employed

		Number	
Self-identified self-employed		**19,667,952**	
Self-employed without employees (own- account)		**11,890,336**	
Self-employed – without employees in an unincorporated enterprise[2]		**11,794,167**	
Self-employed – without employees in an unincorporated enterprise and who do not control the price	**Total**,	**2,012,115**	
	And who get supplies or instructions from another company/agent/contractor (# and per cent of the category)	299,272	14 %
	And who get their own supplies (no instructions) (# and per cent of the category)	1,708,386	85 %

This analysis has indicated that in Thailand, as in other countries where dependent contractor testing has been conducted, the pattern of dependency and indicators of dependency are different in Non-agricultural activities as compared to Agricultural activities (20th ICLS Room document 10). The situation of agricultural producers and conditions of agricultural product markets, particularly regarding pricing, differ from those for non-Agricultural activities.

One third of the workforce in Thailand works in agriculture. More than half (53 per cent) of the self-employed without employees in unincorporated

[2] Among the self-identified self-employed in unincorporated enterprises, 10 percent get supplies from others (1,129,894)

enterprises work in agriculture and 47 percent work in non-Agricultural activities.

In order to assess dependency, three different groups were created based on the two tested boundaries. i.e. self-identified self-employed persons without employees in an unincorporated enterprise who:

(a) Do not control the price and obtain the material input from another company, agent or contractor or customer;

(b) Do control the price and obtain their own material input;

(c) Do not control the price-setting but do obtain their own material input.

Workers in Group a) meet both test conditions for dependent contractor and thus have the highest degree of dependency due to both their economic and their organizational dependency. It is the smallest group (1.7 per cent of self-employed persons without employees in an unincorporated enterprise). With just over one third (35 per cent) of the workers in this group work in agriculture, it has the lowest share of agriculture workers among the three groups The majority of those in this group who work outside agriculture work in their own dwelling.

Workers in Group b) do not meet either condition for dependency; they are likely to be *independent* self-employed workers without employees in unincorporated enterprises. Workers in this group are expected to show the highest degree of independency since they control the price for their goods and services and obtain their own material input. It is the largest of the three group (57 per cent of all self-employed persons without employees in an unincorporated enterprise) and just over half of them work outside of agriculture. The majority of those outside of agriculture are male and 39 per cent work in their own shop or their own dwelling, 25 per cent at a market stall or street and 21 per cent in their own enterprise.

Group c) presents a more challenging case: They procure their own supplies or follow their own instructions, but they report **<u>not</u> controlling price**. They therefore have a degree of economic dependency but a lower degree of organizational dependency compared to group (a). Some 15 per cent of all self-employed persons without employees in an unincorporated enterprise are in group (c). Virtually all work in Agriculture (95%). The 5 per cent who work outside agriculture share characteristics with group (a). The majority of this non-agriculture subgroup work in their own dwelling and an additional 14 per cent have "no fixed" workplace. The ambiguous situation of workers in this group and the high share of agriculture workers point to the usefulness of adding additional boundaries/questions that would assist in interpreting the questions on price control.

Groups (a), (b) and (c) account for 73.3 per cent of the self-identified self-employed persons without employees in an unincorporated enterprise. The remaining 26.7 per cent are not captured by the two test boundaries and hence are not covered by the analysis.

4. Preliminary testing with workers with primary place of work being the home/own-dwelling

A total of 2,557,603 self-identified self-employed without employees in non-Agriculture are home-based workers (includes those in unincorporated and incorporated enterprises):

- of these 2,083,906 or 81% of the total provide their own goods/service; and
- of these 1,395,321 control the price of the goods/services they produce. Another 23,733 negotiate the price with the customer.

Therefore, 1,419,054 workers, or 68 per cent of self-employed workers without employees who are home-based display independency both in operation (get own supplies or instructions) and in pricing (control or negotiate price). Importantly, 268,253 or 13 per cent of home-based self-employed workers without employees could not answer “yes” or “no” to whether they control pricing of their product. Possibly, the pricing question could be improved to address this difficulty. Further work is needed to identify homeworkers – who meet the Dependent Contractor conditions and work from home.

5. Discussion and Conclusion

In conclusion, regarding identifying likely dependent contractors through the *self-employed path*, combining the “price control” criterion with the criterion of getting supplies “from others” provides a good, if narrowly defined, indicator of dependent contractor status for unincorporated own-account self-employed workers, especially those outside of Agriculture. By itself, the price control criterion also is a fairly good indicator. Important next steps would be to examine separately the Agriculture and non-Agriculture workers who do not control price and “get their own supplies.” This is because group c) is overwhelmingly reporting working in Agriculture. This finding in the case of Thailand has also been reported in the trials in other countries (ILO Statistics 2018). We expect that agricultural work mostly entails getting one’s own supplies and receiving no instructions.

With respect to answers to each of these two criteria that indicate likely dependent contractor status we suggest additional questions to be asked of respondents in later surveys. For respondents who answer they do not control price, additional questions can probe who they sell their production to and/or how many customers they have. For respondents who also receive supplies or instructions from others (as well as not controlling prices), particularly for those engaged in agricultural activities, additional questions querying their autonomy in production may yield contextual information to help in determining dependent contractor status.

Additional questions for those who do not control price might include:

- Do you sell your production to one customer? If more than one, how many customers?
- Who do you sell your production to? (Options are: Other farmer/producer; warehouse company; distributor; retailer or other appropriate categories)

Additional questions for those working in agriculture who receive supplies or instructions from others and do not control price, and could be Dependent Contractors include:

- Do you choose the specific variety of the agricultural product you grow (e.g. variety of rice, wheat, legume)? Or does the person or entity that provides supplies or instructions choose the variety you produce?
- Another possible question is: Do you decide how much of your land to cultivate for [specific crop]?

Such contextual variables would enable a more accurate characterization of dependency of own-account operators in Agriculture. A further improvement to explore may be to split the question about "receiving supplies or instructions from another company/agent/contractor or customer" into two separate questions: A questions about "receiving supplies" and another about "receiving instructions." This may elicit further clarification on the type of operational dependency that is relevant in different types of work activity, particularly for comparing Non-agricultural and Agricultural work.

References

1. Carré, Françoise, Michael Frosch, Wissanee Poonsab, and Joann Vanek, (2018). "Thailand 2017 LFS – Highlights of findings from testing of Dependent Contractors," August.
2. ILO Statistics, (2018). 20th ICLS Room Document 10. Testing of proposals for a revised International Classification of Status in Employment. Geneva, Switzerland, October 10-19.
3. 20th ICLS. (2018) Resolution Concerning Statistics on Work Relationships. October 17. https://www.ilo.org/global/statistics-and-databases/meetings-and-events/international-conference-of-labour-statisticians/20/WCMS 647343/lang--en/index.htm

A view on 50 years of life of the ISI: (With a focus on ISI relations with official statistics)

Jean-Louis BODIN
Statistical Consultant, Inspecteur Général de INSEE Honoraire, Paris, France

Abstract

The author is an official statistician. The first ISI Congress he attended was the 37th Session in London in 1969 and he missed only two congresses during these 50 years. The paper is based on his participation to the life of the ISI during these 50 years and mostly focuses on its role in the domain of official statistics.

When ISI was created at the end of the 19th century, most of its objectives were oriented towards what is called today official statistics: international comparisons, development of classifications, population censuses. This role played by ISI changed when the League of Nations (SDN) was created in 1920 and especially after the creation of the United Nations in 1946, when a statistical division was created at the UN Headquarters. So ISI had to find specific ways to continue to have official statisticians as members. The paper will present some issues that have marked the life of the ISI for the past few decades:

- Creation of some sections (now associations) dealing with some aspects of official statistics: IASS in 1973 (surveys), IAOS in 1985 and Irving Fisher Committee in 1997 (central banks statistics)
- Declaration on professional ethics in 1985 and creation of the Advisory Board on Ethics in 2010. Defense of the integrity of statistics.
- Contribution of the ISI to drafting the UN Declaration on Fundamental Principles of Official Statistics in 1991.
- Strengthening statistical capacities in developing countries.
- Role of statistics in the Data Revolution

Keywords

History; official statistics; professional ethics; strengthening capacities

1. Introduction

This paper does not claim to be a history of statistics, nor even a detailed history[1] of ISI. It is only the testimony of an official statistician who has

---

[1] See some historical references at the end of the paper [1] [2] [3]

participated in the activities and life of our Institute since its 37th session[2][3] held in London in 1969 and has contributed more or less actively to all our congresses since the 40th session held in Warsaw in 1975. We are in 2019 and the years ending in 9 have all been especially important for the author: 2019 is the fiftieth anniversary of his first participation in a session of the ISI, the fortieth anniversary of his election as an elected member of the ISI, the thirtieth anniversary of the beginning of his term as IAOS President and the twentieth anniversary of the beginning of his term as President of the ISI.

This paper mostly focuses on role of ISI in the domain of official statistics. It will first describe the main events that occurred during the historical relations between official statistics and ISI until 1969. Then it will present some issues that have marked the life of the ISI for the past five decades in the domain of official statistics.

2. A brief history of the ISI and its relations with official statistics [1] [2] [3]:

The ISI was formally established in June 1885 during the Jubilee of the Statistical Society of London and the 25th anniversary of the Paris Statistical Society with the participation of about twenty countries, almost all European and North-American. Prof. Franz von Neumann-Spallart (Austria) presented on this occasion a paper sketching a plan for an international statistical association. But in fact this creation was only a logical continuation of the holding of the nine International Statistical Congresses (ISC) organized between 1853 and 1876[4] in European cities. The first of these conferences was held in Brussels following an initiative taken by Adolphe Quételet (Belgium) and Prince Albert, consort of Queen Victoria, during the Universal Exhibition organized in London two years earlier. When ISC and ISI were created, most of their objectives were oriented towards what is called today official statistics: international comparisons, development of classifications, population censuses... In his opening address of the 4th ISC held in London in 1860 Prince Albert stated that "*these congresses pave the way to an agreement among different governments and nations to follow up common inquiries, in a*

[2] The name of the World Statistical Congresses of the ISI were "ISI Sessions" until the revision of the Statutes adopted during the 58th ISI session held in Dublin in 2011. The 59th ISI session held in Hong-Kong in 2013 was therefore the first WSC of the new era.

[3] The 37th ISI Session was inaugurated by Sir Harold Wilson, UK Prime Minister. During his speech, he reminded that he was Director of Economics and Statistics at the Ministry of Fuel and Power in 1943–1944 before his first election at the Commons (1946). Later, he was elected as the President of the Royal Statistical Society (1972-73).

[4] No congress was organized between 1876 and 1885 because Germany and Prussia, which were not happy with the way in which the Permanent Commission of the ISC managed them, refused to participate in any international statistical conference. That was probably more a political disagreement than a technical one.

common spirit, by a common method for a common end". This statement is still today a strong recommendation of all international and supranational organisations to their country members. One of the first tasks of the newly created ISI was the production of an *International Statistical Directory*, which is mainly today into the hands of international agencies such as the UN, the World Bank or regional development banks.

Another important achievement of the ISC/ISI before the 1st World War was the production of statistical classifications. The oldest classification to have come into existence was *the International Classification of Diseases and Causes of Death* [4]. First discussions about this classification were made during the 1st ISC in Brussels in 1853 among the different subjects that could be candidates for international statistical comparisons. The 4th ISI session in Vienna (1891) marked the beginning of true international acceptance of statistical lists of causes of death and sickness. Jacques Bertillon, Chief of Statistics for the City of Paris[5], was asked to chair a committee that would prepare a list for causes of death at the next ISI meeting; this classification was adopted in 1893 during the 5th ISI session in Chicago. Early in the history of this classification, a revision cycle was established to keep the list abreast of medical progress. In 1899, during its 7th session in Christiana (former name of the city of Oslo) ISI approved the proposal made by American Public Health Association for a decennial revision of the list. Nowadays maintenance and revision of this classification are made by the World Health Organisation (WHO) based in Geneva.

An interesting controversy took place in the last decade or the 19th century about "representative methods". This controversy followed a paper presented by Nicolai Kiær, director of the Norwegian Central Bureau of Statistics, during the 5th ISI session in Bern (1895): *Observations and Experiments concerning Representative Counts*[6]*.* The reactions of the ISI General Assembly against Kiær's paper were violent and Kiær's proposals were refused almost unanimously. Kiær reiterated his proposals during the 6th session in St. Petersbourg in 1897 and the 8th session in Budapest in 1901 with the same results despite support of Scandinavian and French delegates. After 1903, sampling methods were excluded of the agenda of the ISI sessions until the 16th session in Rome (1925) where the ISI General Assembly approved the conclusions of a committee (Adolphe Jensen, Corrado Gini and Lucien March were among the members of this committee) in favour of these methods.

---

[5] Jacques Bertillon was the grandson of Achille Guillard who investigated the decision to prepare a uniform disease classification during the 1st ISC.

[6] The original version of this paper was presented in French under the title Observations et expériences concernant des dénombrements représentatifs. French and German were at that time the two main languages of the ISI.

Another controversy arose during 13th ISI session in The Hague when the General Assembly decided to create a Permanent Office, precisely in this city. An ad hoc Committee was appointed and its rapporteur, the French Lucien March, wrote that *our institute cannot remain an exclusively academic institution, while the movement of ideas and activities gives rise to ever more pressing needs for statistical information. Some participants announced "the suicide of our Society*" since they feared that "*the setting-up of a permanent office would limit the independence of the ISI*" (it was expected that the operating costs of the permanent office would be covered by government contributions). Fortunately, the existence of this office prevented the ISI from exploding during the First World War: it was located in the Netherlands which was a neutral country, but the President and one vice-president were from allied countries while one other vice-president was German.

This role played by ISI could have changed when the League of Nations[7] (SDN) was set up after the signing of the Treaty of Peace in Versailles in 1919. One of its first acts was to appoint an International Statistical Commission which met in Paris in October 1920 with significant ISI participation[8]. This collaboration gave rise to difficult discussions; some ISI members believed that a learned society should not have to work with an intergovernmental organization on pain of losing its scientific freedom. Another point of view, expressed by Willcox, ISI Vice-president from 1923 to 1947, was that during this period the role of the ISI was that of a "semi-governmental" body. My impression is that these problems have weakened the role of the ISI vis-à-vis the official statistics between the two World Wars, but without compromising its role as a forum for research in probability and mathematical statistics (e.g. the paper on theory of tests presented by Jerzy Neyman during the 22nd ISI session in London in 1934). The outbreak of the 2nd Word War (1939) created a new crisis in the affairs of the ISI. The 24th session convened in Prague in September 1938 closed prematurely because of Hitler's ultimatum sent to the Czechoslovak government. The occupation of the Netherlands by the German forces put an end to all ISI activities; fortunately, the ISI office and its archives were transferred to the Peace Palace which was an exterritorial place.

After the creation of the United Nations in October 1945, it was decided to create the UN Statistical Division (UNSD) at the UN Headquarters. Otherwise, the US government had invited five years ago the ISI to hold its 25th session in Washington, DC, in 1940. Of course it was not possible to organize it. But the Arrangements Committee set up in 1939 for that session merged with the UN, the ISI and other bodies into an Arrangements Committee for International

[7] The League of Nations is best known by its French acronym: SDN (for Société des Nations)

[8] This Commission was chaired by the ISI President Luigi Bodio (Italy) and its Vice-president Albert Delatour (France).

Statistical Conferences. Prof. Stuart Arthur Rice played a fantastic role in this committee; he played a leading role in reactivating[9] the ISI from 1945 onwards, in particular in drafting new statutes and his invaluable efforts were recognized by his election as ISI President in 1947[10]. At the same time, he chaired the UN nuclear session held in October 1946 at Hunter College in New York which made recommendations for the composition and terms of an UN Statistical Commission[11] (UNSC). According to J. W. Nixon, three names have stood out the history of the ISI, Adolphe Quételet with the foundation of the ISC in 1853, Franz von Neumann-Spallart who drafted the first ISI Statutes in 1885 and Stuart Arthur Rice

Influencing governmental statistical agencies by encouraging uniformity in statistical definitions and data collection was certainly one of the main roles of the ISI before the 1st Word war, but after 1945, this role was taken over by the UN. The mission of the ISI became international communication among statisticians rather than among governments, international promotion and development of research, statistical capacity and good practice across all statistical domains. ISI has had consultative status with the UN Economic and Social Council since 1949 and therefore with the UN Statistical Commission. Similar links may be identified with regional and specialized UN bodies, such as the Conference of European Statisticians which is a subsidiary body of the UN Economic Commission for Europe in Geneva, FAO, ILO, UNESCO and others.

After Stuart A. Rice, many ISI leaders were not only involved with the ISI as Presidents or other leading role, including within the ISI associations/sections, but also chaired the UNSC or the Conference of European Statisticians, e.g. Petter Jacoj Bjerve (Norway), Prasanta Chandra Mahalanobis (India), Rao (India), Mikhail Antonovitch Korolev (USSR), Vera Nyitrai (Hungary), Edmond Malinvaud (France), or more recently William Begeer (the Netherlands), Carlos Jarque (Mexico), or Shigeru Kawasaki (Japan). Sir Harry Campion(UK) is another example: after helping setting up the UNSD and was its first head, he also chaired the UNSC in 1951 and 1952 and was ISI President from 1963 to 1967.

[9] Stuart A. Rice also initiated in 1940 the Inter American Statistical Institute (IASI) which still exists today and is the most active regional society affiliated to the ISI.

[10] Stuart A. Rice was re-elected ISI President in 1949 and 1951 and elected Honorary President in 1953

[11] The UN Statistical Commission held its 50th session this year (2019) at the UN Headquarters in New York

3. Presentation of some issues concerning official statistics that have marked the life of the ISI for the past five decades

This chapter does not purport to comprehensively cover all areas of intercourse between the ISI and official statistics, but illustrates some of these interactions.

3.1. The ISI General Assembly convened during the 41st ISI Session in Delhi, India, in 1977, established a Committee on Future Directions. This committee, chaired by Joseph Duncan[12] presented its report [5] during the 42nd session in Manila, Philippines. They made the main following recommendations:

i. The fundamental mission of the ISI will be to serve as the premier international academy of statisticians with a special focus on asserting a highly professional and moral tone on statistical issues of wide-spread public interest and in promoting the development of the professions.
ii. The ISI of the future will regard the integration of statistics as one of its major objectives[13]
iii. The ISI will create search committees in an effort to extend its membership to the developing countries end thereby broaden the geographic representation of its membership.
iv. The professional competence of the ISI membership will be utilized in service to international organizations, national governments and others through an organized program of research and development and consultation, and the creation of the International Center for Research and Development in Statistics.
v. Statistical education and training will continue to be a major role of the ISI in the future: the major focus of such activity will shift from the developed to the developing countries.
vi. ISI will undertake regional programs to expend its influence in the developing regions of the world and to recognize the specific needs of the individual regions.
vii. ISI will seek to strengthen its sections and integrate their programs into the whole fabric of the Institute.
viii. ISI will seek to strengthen its relationship to national statistical societies and ensure that there is a minimum of overlap or duplication in the functions of professional societies at the national and international level.

[12] Joe Duncan was in 1977 – 1979 director of the Office of Federal Statistical Policy and Standards, at the US Department of Commerce.
[13] An ad-hoc committee on the integration of statistics was established during the 40th session in Warsaw in 1975 and presented its conclusions in the same time as those of the Committee on Future Directions in Manila.

These eight recommendations (among over 85 specific recommendations) are still topical forty years after their presentation.

3.1. The decade between 1969 and 1979 that ended by the presentation of these recommendations in Manila has seen important changes in the ISI governance:

- After 1975, it was decided that the term of ISI Presidents will be limited to two years, but the Presidents will be elected two years before the beginning of their term and serve as President Elect; an Executive Committee and a Council replaced the Bureau.
- The creation of sections (now associations) was encouraged; among them sections dealing with some aspects of official statistics: in 1973 International Association of Survey Statisticians (IASS), in 1985 International Associations for Official Statistics (IAOS) and in 1995 the Irving Fisher Committee on Central Banks Statistics. The Presidents of sections may participate in ISI EC and Council meetings as observers.

Before the creation of IAOS, there were other ISI sections dealing more or less with official statistics, for instance, IASS or IASC (International Association for Statistical Computing) Statisticians). These other sections are grouping statisticians interested by the development of methods in their domain. The objective of IAOS was different: several people insisted on giving to this new section the name of International Association **for** Official Statistics and not International Association **of** Official Statisticians. That means that we hoped that IAOS would be open not only to official statisticians but also to the users of official statistics and more generally to all those who are interested by the role of statistics in the society. These goals are still very relevant after 35 years. I was one of the founding persons of IAOS with Vera Nyitray, President in 1985 of the Központi Statisztikai Hivatal (KSH – Central Bureau of Statistics); she was also happy to create an independent forum where statisticians coming from both sides of the Iron Curtain may discuss in a neutral place and a scientific atmosphere.

3.2. In 1979, it was decided to build a set of ethical principles and good practices and a committee in charge of preparing an ISI Declaration on Professional Ethics [6] was established. One of the first attempts to formalise ethical rules was certainly the ASA Code of Conduct for Statisticians adopted in 1979 by the American Statistical Association under the leadership of W. Edwards Deming (now, after a revision in 1989, the ASA Ethical Guidelines for Statistical Practice). W. E. Deming was also, with Roger Jowell (UK), one of the promoters of the ISI Committee. These two codes are very general and apply not only to official statisticians but also to the statistical community at large (academic statisticians, researchers, statisticians working in industry, etc.) and don't take into account three specific characteristics of official statistics:

- Official statisticians do not work for a specific consumer or a small group of users; they receive public funds to be at the service of the society at large and to contribute in their domain to the Citizens' Right to Information.
- The individuals' Right to Privacy very often conflicts with the society's Right to Information (in order to know its collective characteristics).
- The bodies responsible for official statistics have a dual authority, a scientific authority and an administrative authority.

During the preparation of the declaration, a divergence arose in particular concerning the mandatory nature of surveys, which is often the case in official surveys, and on the concept of 'informed consent'.

The ISI Declaration was adopted during the Centenary ISI Session (45^{th} session) held in Amsterdam in 1985. An ISI Committee on Ethics was established but was not very active until 2000. It was reactivated following the holding in Buenos Aires in June, 1998, of an international conference intended to bring the support of the international community to INDEC (Instituto Nacional de Estadística y Censos) after the strong attacks of the President of the Argentine Nation on employment surveys. After careful work by this committee, a revised and modernized version of the ISI Declaration on Professional Ethics was approved by the ISI Council in July 2010 and formally presented on World Statistics Day on 20 October 2010. The same year the committee's activities were terminated and an ISI Advisory Board on Ethics [7] (ABE) was established to advise the ISI EC and Council on relevant ethical issues and to recommends or undertake activities for promoting observance of ethical principles in statistics. Most of the work of the ABE since its creation concerned official statistics (e.g. manipulation of the CPI in Argentina between 2008 and 2015, or deletion of the long form in the Canadian census in 2010). In addition, the ABE also intervened to defend statisticians unfairly prosecuted while they had done their work properly; the iconic cases were those of Graciela Bevacqua (Argentine) and Andreas Georgiou (Greece).

3.4. The fall of the Berlin Wall and the collapse of the communist systems in 1989 had huge consequences; the market-oriented system obliged a far greater number of people with economic and social responsibilities in society to take decisions. Such decisions implied the use of an adequate information system and, in particular, sound and relevant statistical information. Moreover, it was vital for statisticians to gain the confidence of the public in the information they were to produce. In the early months of 1990, statisticians from Central and Eastern Europe were fully aware that it was not so easy to face this new challenge and to gain this indispensable trust of the public. They were looking for new references, new landmarks, to serve as a framework in which to fulfill their duties. Such references were maybe not so different from the strictly technical point of view, but totally different where the concept itself of the role of statisticians in the society was concerned. The reaction of the statistical community to this concern was very rapid. The Bureau of the UN Conference of European Statisticians (CES) decided to organize a consultation and a workshop during the first semester of 1990 [8]. The ISI appointed a representative to the working group[14] that was created to prepare a document to be submitted to the CES. The terms of reference of the working group asked it to take into account «*in particular the work done by the ISI on ethical guidelines for statisticians*". The result of this work was the "Fundamental Principles of Official Statistics" that was endorsed by the 39th session of the CES in April 1992 then adopted by one of the Conference's parent body, the UN Economic Commission for Europe, during its 47th plenary session in April of 1992. Shortly after, it was recognized that these Principles have a universal value and a global significance. Following an international consultation process, a milestone in the history of international statistics was reached when the UN Statistical Commission (UNSC) at its Special Session in April 1994 adopted the very same set of principles – with a revised preamble – as the UN Fundamental Principles of Official Statistics. At its 42nd session in 2011, the UNSC acknowledged that the Principles were still as relevant today as they had been in the past and that no revision of the ten Principles themselves was necessary. The Commission recommended, however, updating the preamble in order to take into account new developments since 1994 and adopted the revised preamble at its 44th session in 2013. In July 2013 the UN Economic and Social Council adopted the Principles on UNSC recommendation and, pursuant to its recommendation, the UN General Assembly, in its resolution 68/261 of 29 January 2014, endorsed the Fundamental Principles of Official Statistics [9]. The ISI played an important role in the preparation, dissemination and popularization of the Principles that are an important work tool for the Committee on Professional Ethics, then for the Advisory Ethical Board when instructing specific files. An important number of papers during ISI sessions, then ISI World Statistical Congresses, were prepared and presented.

3.5. Official statisticians have long been more than reluctant to engage into the measurement of "sensitive political issues" [10]. For a long time, they considered that involving statistics in human rights or democratic governance assessments would endanger their "scientific neutrality". This

[14] In addition to the ISI representative, there were representatives of Poland (chair), France (vice chair), Bulgaria, Romania, Spain, Switzerland and Turkey.

reluctance had also been linked to the quite rudimentary basis of the methods used by human rights activists. The first significant impulse for involving official statisticians in human rights issues was given by the ASA with the creation in late 1970s of its Standing Committee on Scientific Freedom and Human Rights as a reaction to the disappearance and assassination of Carlos Noriega, director of he INDEC, by the Argentinean military dictatorship. Most of the founding members of this ASA Committee expected an enlargement of its action towards exploring possible applications of statistics to monitor the status of human rights; nevertheless they were refrained from engaging the Committee in this path until 1985. At the global level, the ISI started by opposing strong resistance to attempts by some of its members to involve in more human right. In 1981, since the 43rd ISI session was organized in Buenos Aires, the Prof. Edmond Malinvaud, President of the ISI, expressed ISI members' concerns about Carlos Noriega's disappearance to the President of the Argentinean Nation through a visit to the Argentinean Ambassador in Paris. Four years later, a contributed paper meeting on *Statistics, Statisticians and Human Rights*, was initially accepted and scheduled for the 1985 ISI session in Amsterdam, but the ISI EC later decided to remove it from the official program and it was held as an informal gathering. During years, this reluctance to mix scientific statistical work with "sensitive political issues" was broadly shared within the ISI. It was only in 1999, IAOS decided that its 2000 Independent Conference be devoted to the use of statistics to assess human rights and some aspects of democratic governance, in particular in the context of development processes. Nowadays such works have largely been done within ISI, IAOS and the statistical community at large.

3.6. Since a long time, the ISI has been actively helping to build statistical capacity in developing countries. The ISI EC produced in 2013 a so-called White Paper on Statistical Capacity Building (SPC). ISI has been proactive in organizing several workshops on leadership and management in different regions (Anglophone Africa in Addis Ababa, Ethiopia, in 2011 and Dar El Salaam, Tanzania, in 2015; Francophone Africa in Dakar, Senegal, in 2012 and Yaoundé, Cameroon, in 2016; Eastern Asia in Daejon, Korea, in 2012). ISI also organized topical workshops, e.g. on poverty in Kathmandu, Nepal, in 2017 or I sponsoring or co-sponsoring national activities. The ISI Committee on SPC is also one of the partners for the organization of the UN World Data Forum that is a suitable platform for intensifying cooperation with various professional groups, such as information technology, geospatial information managers, data scientists, and users, as well as civil society stakeholders. Two Forums were already organized in Cape Town, South Africa, in 2016, then in Dubai, United Arab Emirates, in October 2018 in Dubai.

4. Conclusion

The ISI has proven through its long history of more than a century and a half[15] that, among all the fields of theoretical statistics and applied statistics, it could also bring the experience of all its members to the development of the official statistics, in all regions of the world, and in particular in developing countries, without jeopardizing its other activities.

References

1. J. W. Nixon A History of the ISI 1885-1960 ISI publications, The Hague 1960
2. History of the ISI - ISI Website: https://www.isi-web.org/images/WSC/2013/ISI%20History%20Brochure.pdf
3. J.-J. Droesbeke & Ph. Tassi Histoire de la Statistique, Collection Que Sais-je ? PUF, Paris 1990
4. History of the Classification of Diseases and Causes of Death National Center for Health Statistics, Center for Disease Control and Prevention, USA 2011
5. Report of the ISI Committee on Future Directions, ISI Permanent Office, October 1979
6. ISI Declaration on Professional Ethics, 1985, revised 2010. ISI Website: https://www.isi-web.org/index.php/news-from-isi/150-ethics-intro
7. The ISI Advisory Board on Ethics, 2010 ISI Website: https://www.isi-web.org/index.php/activities/professional-ethics/isi-advisory-board
8. J.-L. Bodin, The Fundamental Principles of Official Statistics, the Breakthrough of a New Era, Chapter 5 of the book "50 years of the Conference odf European Sttaisticians", UN Geneva, 2003
9. Resolution on the Fundamental Principles of Official Statistics, with revised preamble 2013. UN Website: https://unstats.un.org/unsd/dnss/gp/FP-Rev2013-E.pdf
10. Measuring Human Rights and Democratic Governance, OECD Journal on Development, Vol. 9, No 2, 2008.

[15] 134 years since the creation of the ISI and 166 years since the 1st International Statistical Congress in 1853

Steve O. Rice (1907 – 1986) – Inspired by random noise, inspiring statistics research

Georg Lindgren
Mathematical statistics, Lund University, Sweden

Abstract
"In 1944 and 1945, S. O. Rice published a monumental study of noise, generally regarded to be the single most useful source of information about Gaussian noise." These strong words (from the history of research at Bell Telephone System) illustrate how abstract theoretical concepts, like Gaussian processes, correlation function and power spectrum, became common goods in one, very specialized, field of engineering, communication theory. The title of the papers was ``Mathematical analysis of random noise". Steve Rice is regarded as a giant in communication theory. He deserves a similar place as a door opener for the statistics community to a rich field of modern important problems, in reliability, environmental research, transportation, and other fields.

Keywords
Extreme value theory; Rice's formula; stationary processes; Gaussian process; level crossings

1. Introduction

Stephen O. "Steve" Rice is known in probability theory mainly through "Rice's formula" for the average number of level crossings per time unit by a stationary stochastic process. However, his influence on statistics research goes far beyond level crossings. Steve Rice was a member of the Bell Telephone Laboratories, USA, from 1930 to his retirement 1972, working as a consultant on transmission engineering and doing research in communication theory. Stochastic variation penetrated all his work and he mastered to combine the mathematical analysis with the concrete questions at hand. He was well aware of the theoretical literature on stochastic processes that emerged during the 1930s but argued that the work by Khintchine and Cramer on Fourier representation of´ the covariance function, lacked contact with important concrete engineering problems. His two papers from 1944 and 1945 on "Mathematical analysis of random noise"[13] delivered that contact.

Many engineering communities quickly applied the stochastic approach to previously deterministic fields, and control, mechanical, and ocean engineering formulated their own versions, based on Rice's treaties. One spectacular shift in paradigm took place in ocean engineering, which built advanced stochastic models for the ocean surface already in the early fifties, negating the quote by Lord Rayleigh: "The basic law of the seaway is the apparent lack of any law". Central in many engineering studies were quantities like maximum and local maxima of random processes, number of level crossings and the distance between them. Rice's formula for the expected number of level crossings was used extensively.

The statistics community was slow to appreciate the many challenges that lay open in Rice's two articles, but around 1960 theoretical studies on level crossings and extremes started to appear in core statistical journals and conference proceedings. The book "Stationary and related stochastic processes, by Harald Cramer and Ross Leadbetter (1967),[2] set the style for much of theoretical research. It also gave a firm basis for the coming development of statistical extreme value theory and its applications.

2. The person and his career

Steve O. Rice was born on November 29, 1907, in Shedds, Oregon, USA, as the only child of Stephen Rice, a buttermaker, and Selma R. Bergren.[*+] Steve entered Oregon State University, Corvallis, and received a B.S. degree in electrical engineering in 1929. After a year of graduate studies in physics at California Institute of Technology, Pasadena, he joined Bell Telephone Laboratories in New York, 1930. With Bell he got freedom to do own research on the mathematical background of communication systems, but he also earned a position as a very knowledgeable consultant for different communication technology groups. During many decades Bell Labs was a hot-bed for the development of communication theory and technology, to mention just two names with statistics interests besides Steve Rice: Claude Shannon with Information theory, and David Slepian with "Slepian's comparison lemma" and Coding theory. When Rice retired in 1972 he was "Head, Communication Theory Department", and located in Murray Hill, New Jersey.

Steve Rice's deep knowledge in mathematics and stochastic processes was an invaluable asset, not the least since he understood how to combine the

---

[*] Some details in this section are extracted from David Slepian's "Memorial Tributes", 1991.[16]

[+] Selma Bergren's father was born in Sweden and her mother had a typical Swedish name, while born in USA theory with the different applications. Many are those who have witnessed about his kindness and helpfulness, to young PhD students as well as to established researchers.

3. An overview of his work

Steve Rice published sixty-four scientific papers during his career. In the list[8] they are categorized into three groups according to focus: Computation, mathematics, and statistics; Physics and communication systems; Signal processing.

"Mathematical analysis of random noise",[13] with its 162 pages, belongs to all three categories. Its significance for radio communication and general signal processing is obvious – it was written for engineers in an engineering environment. Rice was quite familiar with the work by Norbert Wiener on correlation functions, harmonic analysis, and filtering of certain "random" functions, written as mathematics with electrical engineering applications in mind. Rice embraced explicitly the idea of a "stationary stochastic process" as an ensemble of functions with a statistical distribution in the sense of Khintchine and Cramer. Wiener, as prestigious mathematician, helped to advocate the statistical´ viewpoint, when he claimed that information is not only what has been said, but also what might have been said. It has been argued that Rice's "Random noise",[13] Shannon's "Information theory",[15] and Wiener's book[18] on smoothing and prediction, are the three most important publications from USA during the 1940s to establish the statistical radio communication theory.

"Mathematical analysis of random noise" has four main parts: (I) on shot noise, (II) on power spectra and correlation, (III) on the statistical properties of correlated noise, (IV) on non-linear filters. Part
(II) presented a stationary process as a random Fourier sum, with discrete spectrum,

$$x(t) = \sum_n a_n \cos \omega_n t + b_n \sin \omega_n t = \sum_n c_n \cos(\omega_n t - \phi_n), \tag{1}$$

with random (a_n, b_n) and (c_n, ϕ_n). In part (III) the discrete spectrum is replaced by a continuous one and the signal $x(t)$ assumed to be Gaussian.

It took only a decade before the statistical content of parts (I, II, IV) was included in advanced textbooks for radio engineers.[3] But what makes Rice's random noise paper important for statistics research is part (III), and the statistics community was slow to recognize its challenges, namely the statistical properties of the number and location of level crossings by a stationary process. We will deal with some of these challenges later in this paper.

Among Steve Rice's many deep contributions to communication theory, David Slepian[16] mentions three as particularly influential. His 1950 paper on "Communication in the presence of noise" was the first to evaluate explicit bounds for the error probability in information systems. "Reflection of electromagnetic waves from slightly rough surfaces", from 1951, was fundamental for the understanding of radar returns from the ocean. It was an early example of the Fourier model (1) generalized to space-time. The third example given by Slepian is the 1963 analysis of "Noise in FM receivers", an ingenious analysis of a mysterious phenomenon.

4. Rice's formula and its ramifications

The modern version of Rice's formula for the expected number of upcrossings of a given level u per time unit (="rate of upcrossings") by a differentiable stationary process reads

$$\mu_u^+ = E(\#\{t \in [0,1]; X(t) = u, upcrossing\}) = f_{X(0)}(u) E(X'(0)^+ | X(0) = u) \quad (2)$$

We'll discuss this formula for a stationary processes with spectral density $S(\omega)$ and covariance function $r(t) = \int \exp(i\omega t) S(\omega) d\omega$, and spectral moments $\lambda_{2n} = \int \omega^{2n} S(\omega) d\omega$.

a) The origin of Rice's formula[11] for the rate of level crossings motivates the title of the present paper – "inspired by random noise, inspiring statistics research". The root of the formula is in a study[12] from 1939 on the distribution of the height of local maxima of a random function, stating: "If suitable conditions are satisfied the probability that the random curve $y = F(x)$ has a maximum in the rectangle $(x_0, x_0 + \mathrm{dx}_0; \mathrm{y}_0, \mathrm{y}_0 + \mathrm{dy}_0)[\dots]$ is $p(x_0, y_0)\mathrm{dx}_0\mathrm{dy}_0$ where $p(x_0, y_0) = -\int_{-\infty}^{0} P(y_0, 0, \zeta)\zeta d\zeta$." Here $P(y_0, 0, \zeta)$ is the joint density of the curve and its first and second derivative, taken at the point $x = x_0$.

In the paper Rice motivates his interest: "Here the distribution of the maxima of [...] random curves is studied. Although this problem is of some physical interest I have been unable to find references to any earlier work. Problems of this nature occur in the investigation of the current reflected by small random irregularities along telephone transmission lines."

Integrating out y_0 in the maximum formula Rice obtains the rate of local maxima regardless of height, i.e. the rate of zero downcrossings by the derivative of the process. He also rewrites the formula to give the average number of downcrossings. With modern notation his formula reads,

$$E(\#\{t \in [0,1]; X'(t) = 0, \text{downcrossing}\}) = \int_0^1 f_{X'(s)}(0)\, E(X''(s)^- | X'(s) = 0)\mathrm{ds}. \quad (3)$$

Thus, Rice's formula for the average number of zero (up- or down-) crossings is from 1939.[12]

In "Random noise" Rice arrives at (3) by probabilistic arguments with a stationary processes in mind. He also mentions that "equivalent results [on the expected number of zeros] have been obtained independently by M. Kac".[5] However, the context of the Kac formula is strictly the number of roots to an algebraic equation, chosen at random under a variety of explicit assumptions about its distribution. Due to the technical similarity of the two formulas, the term "Kac-Rice" formula" is sometimes used.

b) Strict proofs, alternatives, and higher moments were late to be developed, and not until the early1960s definite conditions for the validity were published. The book by Cramer and Leadbetter from´ 1967[2] gives an accurate account of the knowledge in the mid 1960s.

Consider first a stationary Gaussian process with mean zero. Then $\mu_u^+ = \frac{1}{2\pi}\sqrt{\lambda_2/\lambda_0 e^{-u^2/2\lambda_0}}$ is easily obtained from (2), with $\lambda_0 = \mathsf{V}\big(X(t)\big), \lambda_2 = \mathsf{V}\big(X'(t)\big)$. When the process is non-differentiable $\lambda_2 = \infty$. Ylvisaker[19] proved 1965 that the formula for μ_u^+ holds, both for finite and for infinite λ_2.

For the variance it took even longer and not until 1972 the final detail was filled in by Don Geman.[4] For the variance of the number of zeros of a Gaussian process to be finite it is necessary (Geman) and sufficient (Cramer and Leadbetter[2]) that $\int_0^\delta (r''(t) - r''(0))/t \, dt \; < \infty$ for some $\delta > 0$. In fact, the result holds for any level.

c) "The problem of determining the distribution function for the distance between two successive zerosseems to be quite difficult and apparently nobody has as yet given a satisfactory solution". This quote from Section 3.4 of "Random noise" still carries a grain of truth. The length of excursion interval was the central theme of research following "Random noise" during the first 15 years and it has gained renewed popularity from around 1990 from applications in physics and material science.

For the distribution of the length of an excursion above zero, Rice suggested a Bonferroni type inclusion-exclusion series. The excursion time density $f_T(t)$ approximated with the first three terms in the Rice in- and exclusion series is

$$f_T(t) \approx \frac{1}{\mu_0^+}\left\{\mu_0^{+-}(0,t) - \int_{s=0}^{t} \mu_0^{+-}(0,s,t)ds + \iint\limits_{0<s_1<s_2<t} \mu_0^{+---}(0,s_1,s_2,t)ds_1 ds_2\right\}. \quad (4)$$

Here $\mu_0^{+-}(0,t)$ is the bivariate intensity that there is a zero upcrossing at 0 and a downcrossing at t with no restriction to what happens in between. That term obviously gives an overestimation of the density. Subtracting the integral $\int_{s=0}^{t} \mu_0^{+--}(0,s,t)ds$ that there is an extra downcrossing somewhere $(at\, s)$ gives an underestimation, and so on. Improvements to the moment based Rice series was a recurrent theme in applied studies of excursion times, during several decades.[10]

d) Rice's handling of crossing and multiple crossing intensities was quite intuitive and the strict in-terpretation of "the distribution" of an excursion was by no means clear. For example, one can ask for the conditional probability that $X(s) > 0, 0 < s < t$, given that $X(0) = 0$, upcrossing, an event that has probability zero for any stationary Gaussian process. That would give the probability that the excursion lasts more t time units after an excursion-start at 0.

Kac-Slepian's h.w. (horizontal window) conditioning (1959)[6] gave a precise meaning to the conditioning on the zero-probability crossing event. All conditional probabilities should be defined as the limit of a well defined conditional probability, given a crossing, not exactly at 0, but in a horizontal window $[0,h]$ the limit taken as $h \downarrow 0$. Kac and Slepian also showed, by an ergodic argument, that a h.w. conditioned distribution is equal to the limit of the corresponding empirical distribution observed after all u −upcrossings $t_k > 0$, t_k as the observation interval goes to infinitity: the probability that an excursion exceeds t is

$$P(X(t_0+s) > u; 0 < s < t | X(t_0) = u, \text{h.w upcrossing})$$

$$= \lim_{T\to\infty} \frac{\#\{t_k < T; X(t_k+s) > u, 0 < s < t\}}{\#\{t_k < T\}} = \frac{E(\#\{t_k < 1; X(t_k+s) > u, 0 < s < t\}}{E(\#\{t_k < 1\})}. \quad (5)$$

Thus, the meaning of the condition is clear; what remains is to compute the expectation in the nominator in (5). Rice backed away from the difficult integrals involved in the higher order approximations in the Rice series (4). Interestingly enough, he mentions the possibility to use the condition that $X(t_k + s_j) > u$ for equally spaced points s_j between 0 and t. He concludes that also these integrals should be hard to evaluate. Indeed, advances in statistical computing has made it possible to compute the expectation in (5) with very high accuracy in reasonable time on a standard computer; see Lindgren.[7]

e) The 1967 book by Cramer and Leadbetter on stationary processes[2] made crossing problems, Rice's formula, and its consequences available for the general statistical community. The book also represented a link between the moment based crossing analysis by Rice and others on the stream of crossings, and the emerging extreme value theory for dependent sequences and processes. One of its most important theorems says that under conditions that restrict local clumping and dependence over long distances the stream of level upcrossings form an asymptotic Poisson process as the level increases. The proof given in Cramer and Leadbetter[2] leans heavily on computations in Rice[14] and points forward to modern extremal mixing conditions.

Rice's formula together with the Poisson approximation for high level upcrossings is often used in reliability analysis to approximate the probability of the occurrence of an extreme value in a differentiable stationary load process: P (no upcrossing of level u before time t) $\approx e^{-t\mu_u^+}$.

5. Influence of "Random noise" on stochastic modelling

The comprehensive treatment of random noise in communication systems in "Random noise" spurred stochastic research also in applied physics, acoustics, and material science, but it had remarkably little effect on statistics and probability. A rare exception is M.S. Bartlett's book[1] from 1955 on methods and applications of stochastic processes, focusing on applications. Bartlett mentions Rice, en passant, in connection with the recurrence time of level crossings.

Much more "statistically significant" is the dramatic effect that "Random noise" had on physical oceanography and naval engineering. In the early 1950s statistical studies of ocean waves concentrated on wave height distributions, time correlation, and Fourier analysis. In an unusually fortunate cooperation between Willard Pierson, an oceanographer, and Manley St Denis, a naval architect (ship builder), wrote a pioneering paper on the motion of ship on a stochastic ocean.[17] Heavily inspired by Rice's "Random noise" they developed a full Gaussian model in time and 2D space for irregular waves, generalizing (1) to elementary waves of different wavelengths coming from different directions. They also implemented filtering techniques to describe the movements of a ship sailing on the ocean and derived the necessary filter functions.

Michael Longuet-Higgins was a British oceanographer who published more than twenty articles on statistical properties of waves between 1952 and 1991. A paper from 1957[9] is a parallel to Rice's "Random noise" in its detailed description of stochastic characteristics of a random sea surface. His analysis 1962[10] of Rice's in- and exclusion series for the distribution of intervals between zeros in a Gaussian process stretches the techniques as far as can be expected; further exact results had to wait until numerical algorithms and computer technology had advanced.[7]

Acknowledgement

I want to thank Krzysztof Podgorski and Igor Rychlik for comments on this paper and for many years´ of cooperation on the statistical consequences of Rice's formula.

References

1. M. S. BARTLETT, *An introduction to stochastic processes with special reference to methods and applications*, Cambridge University Press, 1955.
2. H. CRAMER AND´M. LEADBETTER, *Stationary and related stochastic processes*, Wiley, New York, 1967.
3. W. DAVENPORT AND W. ROOT, *An introduction to the theory of random signals and noise*, McGraw-Hill Book Company, New York, 1958.
4. D. GEMAN, *On the variance of the number of zeros of a stationary Gaussian process*, Ann. Math. Statist., 43 (1972), pp. 977–982.
5. M. KAC, *On the average numbert of real roots of a random algebraic equation*, Bull. Amer. Math. Soc., 49 (1943), pp. 314–320.
6. M. KAC AND D. SLEPIAN, *Large excursions of Gaussian processes*, Ann. Math. Statist., 30 (1959), pp. 1215–1228.
7. G. LINDGREN, *Gaussian integrals and Rice series in crossing distributions – to compute the distributions of maxima and other features of Gaussian processes*, Stat. Sci., 34 (2019), pp. 1–31.
8. ____, *Scientific publications by Steve (Stephen) O. Rice (1907-1986)*. . https://lucris.lub.lu.se/admin/files/61678271/RicePublications.pdf, 2019.
9. M. S. LONGUET-HIGGINS, *The statistical analysis of a random, moving surface*, Phil. Trans. Royal Soc. London, A, 249 (1957), pp. 231–387.
10. ____, *The distribution of intervals between zeros of a stationary random function*, Phil. Trans. Royal Soc. London, A, 254 (1962), pp. 557–599.
11. A. RAINAL, *Origin of Rice's formula*, IEEE Trans. Inform. Theory, 34 (1988), pp. 1383–1387.
12. S. RICE, *The distribution of the maxima of a random curve*, Am. J. Math., 61 (1939), pp. 409–416.
13. ____, *Mathematical analysis of random noise*, Bell Syst. Tech. J., 23 (1944), pp. 282–332., 24 (1945), pp. 46–156.
14. ____, *Distribution of the duration of fades in radio transmission: Gaussian noise model*, Bell Syst. Tech. J., 37 (1958), pp. 581–635.
15. C. SHANNON, *A mathematical theory of communication*, Bell Syst. Tech. J., 27 (1948), pp. 379– 423.
16. D. SLEPIAN, *Steven O. Rice 1907–1986*, in Memorial Tributes, vol. 4, National Academy of Engineering, 1991.
17. M. STDENIS AND W. J. PIERSON, *On the motion of ships in confused seas*, Trans. Soc. Nav. Archit. N.Y., 61 (1953), pp. 280–332.

18. N. WIENER, *Extrapolation, interpolation, and smoothing of stationary time series. With engineering applications*, Wiley, New York, 1949.
19. N. D. YLVISAKER, *The expected number of zeros of a stationary Gaussian process*, Ann. Math. Statist., 36 (1965), pp. 1043–1046.

Opportunities and challenges facing China's Official Statistics

Ning Jizhe
Commissioner of the National Bureau of Statistics of China

Abstract

Proceeding from the importance of official statistics on modern state governance, the paper elaborates on the opportunities and challenges facing China's official statistics over recent years and the way forward. The paper holds that China's official statistics is developing on the fast track, facing unprecedented opportunities provided by the demand for modernization of state governance system and governance capacity, the demands for statistical data from all sectors of society, the development and application of modern information network technology, and the ever-growing international exchange and cooperation. Meanwhile, the governments at various levels pose increasingly higher demands upon statistical indicators and statistical data; the respondents are growing in number; and the application of modern statistical approaches is relatively backward. All these constitute new challenges to official statistics. To respond to the new requirements under new circumstances, Chinese official statistical offices will endeavor to build and improve a statistical system conducive to high-quality economic development, deepen the reform on system of national accounts, advance information disclosure, data sharing and statistical analysis and interpretation, organize the construction of statistical cloud infrastructure, strengthen statistical supervision according to law, and implement monitoring on the statistical indicators of the UN 2030 Sustainable Development Goals, so as to make due contributions to the development and progress of official statistics in the world.

Keywords

Opportunities; Challenges; Reform

1. Introduction

Official statistics constitutes an important basis for macro control and informed governance of the country and the ability of collecting and analyzing statistical information represents the governance capability of a modern country. Throughout the world, digital governance has increasingly become the major means of state governance. We always need to refer to authentic, accurate, complete and timely data when it comes to make judgments about the economic situation, develop policies for national macro control, or conduct management upon the sectors concerning social development and people’s livelihood like education, healthcare and poverty alleviation, and employment.

This year marks the 70th anniversary of the founding of the People's Republic of China. Through the development over the past seven decades, a sound and inclusive Chinese official statistical system has basically taken form, which conforms to the national conditions of China, aligns with the international statistical rules, and well meets the needs for economic and social development. **First**, the work of national accounts has witnessed continuous improvement. China’s System of National Accounts (2016) has been developed on the basis of China’s national conditions and the latest international standard, namely, the System of National Accounts 2008. Besides, three major reforms on unified accounting of regional GDP, compilation of national and regional balance sheets, and compilation of natural resource balance sheets are being vigorously advanced. **Second**, the economic statistical survey and monitoring system consisting of regular statistical survey system on the industrial sector, the service sector, the sector of wholesale, retail, lodging and catering, fixed asset investment and real estate development has been established to serve the goal of maintaining economic performance within a proper range by providing timely and accurate data and making analysis, research, judgment and early warning. **Third**, we have established a periodic census system consisting of population census and agricultural census conducted every 10 years, and economic census conducted every 5 years. **Fourth**, we have established a statistical survey system concerning people’s livelihood, which consists of the regular statistical survey programs including agriculture, prices, population, employment and household income and expenditure. **Fifth**, we have established a statistical survey system about ecological environment and sustainable development to conduct statistical monitoring upon energy, resources, environment and addressing climate change; have developed the green development indicators; keep advancing statistical monitoring on the sustainable development goals. **Sixth,** we have established a system of official statistical organizations featuring unified leadership and tiered responsibilities, which consists of the National Bureau of Statistics and local survey offices under its vertical management, statistical agencies of local governments and statistical departments of line ministries. The National Bureau of Statistics leads and coordinates nationwide statistical work, supervises upon its survey offices as well as statistical agencies of local government at various levels, and guides and coordinates the work of the statistical departments of line ministries.

2. Opportunities

In the new era of socialism with Chinese characteristics, China's statistical work has entered a new stage of reform and innovation and embraced unprecedented opportunities.

First, the modernization of state governance system and governance capability provides unprecedented opportunities for official statistics development. Over the past 40-plus years' development since the reform and opening-up, China has grown to be the world's second largest economy; it is in the crucial period of transforming the developing mode, improving the economic structure and changing the growth driver. To gain a whole picture about economic and social development, and promptly detect potential risks and hazards in economic and social development demands authentic, accurate, complete and timely data, which makes the government at all levels pay attention to, need and support statistics all the more, and provides brand new and broad space for official statistics development.

Second, the demand for statistical data from all sectors of society provides unprecedented opportunities for official statistics development. The public service institutions more and more rely on statistical data in management and service delivery; the enterprises have growing needs for statistical data in the process of investment, production and marketing; the experts and scholars have increasing needs of statistical data for analysis and research; and it is more and more the case that the general public can't do without statistical data in their everyday life. All these constitute a mighty impetus for the development of statistical work.

Third, the development and application of modern information network technology provides unprecedented opportunities for official statistics reform. Since the start of the 21st century, global science and technology innovation have entered a period of unprecedentedly intensive activity. Science and technology exert profound impact upon economic and social development as well as people's life and they will have significant impact upon statistical work as well. For instance, the extensive use of modern information technologies like big data, cloud computing, Internet of Things, mobile internet and artificial intelligence generates massive real-time data, which provides a brand-new source of basic data for official statistics. Besides, the rapid development of modern information technology provides advanced techniques and tools for reforming the production mode of statistical data and improving the efficiency of statistical work.

Fourth, the ever-growing international exchange and cooperation provides unprecedented opportunities for official statistics development. For one thing, we can draw on the developed countries' mature experience and practice regarding statistical rules, concepts and standards so as to reduce and avoid detours. For another, China's official statistics has made useful attempts and notable achievements in online statistical reporting, e-accounting and statistics about new industries, new type of businesses and new business models. Therefore, China is contributing "Chinese wisdom" and "Chinese solutions" to the reform and innovation of official statistics of various countries around the world.

3. Challenges

Meanwhile, China faces myriad challenges in official statistics development. China is the most populous, the largest developing country and the largest emerging economy in the world. With a vast territory and uneven development from region to region, China sees complex and changing economic and social organizations, and records high mobility of population. In this context it is increasingly difficult to conduct statistical surveys in China.

First, the governments at various levels have increasingly high demand for statistical indicators and statistical data. Along with the rapid economic and social development, China is seeking high-quality development rather than high-speed growth. It entails the establishment of a new set of statistical indicator system to reflect how the innovative, coordinated, green, open and shared development is pursued, to advance high quality development, to deepen the supply-side structural reform, and to show the status of the "new economy" development. The government at all levels needs rich up-to-date statistical data to conduct precision management of economic and social activities, which requires the statistical survey to consider both the nation's overall needs and the specific governance needs of the local government. Therefore, it is more difficult to do a good job in designing the statistical survey scientifically.

Second, statistical survey has to cover a growing number of respondents. For one thing, the respondents of the statistical survey have more and more complex backgrounds since new industries, new type of businesses and new business models among other forms of the emerging economy and new growth drivers keep emerging. New production modes as reflected in cross-border, integrated and mixed operation spring up in large numbers, with business entities engaging in diversified operations. Against this background, it is difficult to use the old statistical standards to cover all the new economic forms. For another, the respondents care more about their privacy and are protective of commercial secrets; they are more and more reluctant to reveal information about their marital status, income and housing; they dislike to let others know sensitive commercial information related to corporate finance; and they feel more bothered about having to use time and energy as well as related resources for the survey.

Third, the application of modern statistical approaches is relatively backward. Currently, the National Bureau of Statistics of China requests enterprises to fill in and submit data online under the premise of data security, which greatly improves data production efficiency. Given the rapid development of modern information technology, research upon new technologies like big data, cloud computing, Internet of Things in official statistics and their application remain at the initial stage.

4. Reform

Facing the new opportunities and challenges, the National Bureau of Statistics of China has set forth the overarching goal for future statistical reform and development: to form an innovation-driven and law-based statistical work pattern that is more open and transparent; to gradually realize sound and standardized statistical surveys, rigorous and efficient statistical management, inclusive and high quality statistical services, with marked improvement in statistical operation efficiency, data quality and service delivery; and to speed up efforts in building a modernized statistical survey system. For the present and the following period of time, our key tasks are as follows:

First, build a statistical system conducive to high-quality economic development. We have formulated the Implementation Opinions on Building a Statistical System Conducive to High Quality Development, with a view to reflect how the innovative, coordinated, green, open and shared development is pursued. We will continue to improve the statistical monitoring system for new industries, new type of businesses and new business models, as well as the monitoring of online retail and commercial complexes; enrich the statistics on quality and efficiency improvement of the industry; strengthen the statistics about the emerging service industry; further deepen statistical reform of energy, investment and service sectors; carefully implement the monitoring of economic performance on a monthly/quarterly/yearly basis, and properly wrap up the fourth national economic census and prepare for the seventh national population census.

Second, deepen the reform on system of national accounts. Starting from the fourth quarter of 2019, we will carry out unified accounting of regional GDP, speed up efforts to compile national and regional balance sheets, explore the development of balance sheet of natural resources, steadily advance the formulation of the supply and use table, further improve the accounting of capital flow, and the value-added of new industries, new type of businesses and new business models as well as the derivative industries, update the methodology of annual accounting of the residents' self-owned housing services, and promote harmonization with the 2008 System of National Accounts.

Third, advance information disclosure, data sharing and statistical analysis and interpretation. Following the principle of keeping statistics as open as possible, we will step up efforts to make statistics more transparent, improve the information disclosure mechanism, expand the access to micro-data, push for inter-departmental data sharing to eliminate the phenomena of "data island", strengthen analysis and interpretation of statistical data, and guide the public to correctly use the statistics by providing them with rich and detailed data coupled with timely and authoritative interpretation.

Fourth, organize the construction of statistical cloud infrastructure. We will make full use of modern information technology to build a national statistical cloud platform to underpin the operation of statistical work and management of statistical data as well as statistics-related administrative affairs, and promote the application of big data in statistical work.

Fifth, strengthen statistical supervision according to law. We will further improve the statistical laws and regulations, strictly enforce the implementation, strengthen the popularization and publicity, and speed up credit system building in the field of statistics, with a view of enhancing everyone's awareness of statistics production according to law.

Sixth, implement monitoring on the statistical indicators of the UN 2030 Sustainable Development Goals. We will closely follow the latest progress in international work towards the goals set in the 2030 Agenda, conduct in-depth study on the definition and methodology of various indicators in the global indicators framework, and accelerate research on localizing SDG indicators.

China's statistical office has been all along committed to cooperation and communication with statistical agencies of countries and international organizations by engaging in diverse international statistical affairs, learning and drawing on others' outstanding ideas and achievements. We, reformative, open-minded and innovative, are ready to further strengthen multi-lateral, bilateral and regional statistical cooperation with other countries, conduct in-depth communication and mutual learning on statistical theory, system and mechanism, survey methodology, big data application, new economy statistics and capacity building etc., and jointly promote development and progress of official statistics in the world.

ISIWSC2019

Organised by :

DEPARTMENT OF STATISTICS MALAYSIA
MINISTRY OF ECONOMIC AFFAIRS

BANK NEGARA MALAYSIA
CENTRAL BANK OF MALAYSIA

MALAYSIA INSTITUTE
OF STATISTICS

Supported by:

MINISTRY OF TOURISM,
ARTS & CULTURE MALAYSIA

Malaysia Convention & Exhibition Bureau

Visit Truly Asia Malaysia 2020

ISBN 978-967-2000-62-4

#ISIWSC2019