Using Households Surveys to Maintain
the Statistical Business Register

Saleh ALKAFRI, Ph.D.*
Palestinian Central Bureau of Statistics (PCBS), Ramallah, Palestine
saleh@pcbs.gov.ps

Abstract[1]

The sample frame of each annual economic survey (ES) is the latest available Establishment Census (CE), minus the enterprises found closed after the previous ES in Palestine since the sample frame doesn't depend on an Administrative Business Register (ABR). Although attempts have been made in the past to update this frame with new establishments, there is no enough information to do that effectively. Statistics shows that the number of establishments and their employees from 2003 to 2004 and from 2006 to 2007 jumped without any economic reasons. This is due to the fact that in 2004 and 2007, the series has been anchored to the new establishment censuses. In the years before those two censuses, the number of enterprises and their employees was declining, as closed enterprise reduced the sample without newly created units being added. After the 2007 CE, such a decline is not visible, as the economic surveys have been adjusted each year with the developments in enterprises as measured in the quarterly, labour force survey (LFS). This means that each year's economic survey reflects the total number of enterprises in the different activities and the corresponding employment, gross output and value added. This paper introduces the case study of Palestine that mainly utilizes the establishments census to build and update the SBR on a five-year-basis, and the annual establishments and households surveys to update the SBR on an annual basis. This will be a valuable assessment of an actual experiment on how to maintain and update the SBR. The evaluation tools depend on simple comparative calculations. The various calculations show that no selection of LFS data over the period 2007 to 2012 provides a satisfactory projection of the levels of employment and number of enterprises found in the 2012 establishment census. In theory, the methods used should give fair estimates and the tests described didn't show obvious weaknesses in the LFS, but in practice, the 2012 estimates do not match with a substantial margin. It may therefore be expected that use of the LFS to correct the ES, although better than not adjusting the ES at all between censuses, will require further adjustments after conducting a new census.

Keywords: Statistical Business Register (SBR), Sample Frame, Labour Force Survey (LFS), Sample frame maintenance.

1. Introduction

The best practice of the Statistical Business Register (SBR) is to utilize the Administrative Business Registers (ABR) to update it regularly, but if the case doesn't depend on ABR due to under coverage and inaccurate registration in the country, it will be a challenge to maintain, improve and ensure high quality and unit coverage. This paper introduces the case study of Palestine that mainly utilizes the establishments census to build and update the SBR on a five-year-basis, and the annual establishments and households surveys to update the SBR on an annual basis. This case shows how the household labour force survey can be used on the macro level to maintain the SBR in the absence of the ABR. Since there are dramatic changes in the unit of the SBR on daily basis, the census cannot monitor those changes for annual statistical purposes. In addition, the annual establishments surveys just monitor the dead large establishments and the changes in the characteristics of a sample of the establishments on annual basis. Therefore, the outcome of those updates weakens the quality of the SBR, its characteristics and the prospected advantage of being a good source for the sampling frame.

The Labour Force Survey was the solution to adjust the weights of the establishments surveys that are utilizing the SBR as a sampling frame. The paper attempts to review the experience of Palestine on the methods and techniques that were applied over five years, between the two previously implemented censuses (2007-2012), and to evaluate this experience. This experience also can be used for other purposes when the SBR depends on ABR to measure and improve the quality of the SBR.
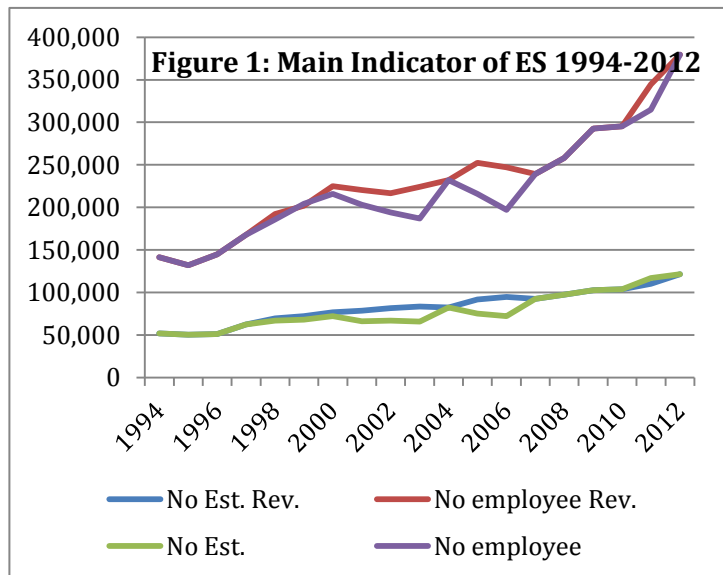
2. Establishment censuses (CE)

The Palestinian National Authority started its partial control on the Palestinian territories in 1993, after Oslo accords. The national statistical office started from scratch. There was no solid available national statistics.

The first CE in Palestine was conducted in 1994. The main use of the census was to produce a sample frame for the Economic Surveys since no business registration was available at that time. The overall annual rate of change, after the initial jump from 1994 to 1997 has been fairly consistent at nearly five percent in terms of businesses and one percent more, approximately six percent, in employment. This overall change does hide a huge turnover in enterprises in existence.

The sequence of censuses was not evenly spread through time, with approximately 3, 7, 3 and 5 years between successive exercises. The changes from one census to the other are therefore skewed. The annual attrition rates mentioned above; however, are corrected for the differences in time between the censuses. It may be noted that the attrition rates for establishments are generally higher than that for employment. This means that the units that close were mostly the small ones. The very last line of the table, the share of businesses and their employment that were established between censuses provides important insights in the turnover. The 2004 census indicated that two thirds of the units had been established since the previous census in 1997. Similarly, 43% of the units found in the 2012 census did not exist in the year 2007.

3. Sample Frame of the Economic Surveys

The economic surveys are based on a sample frame of establishments that is obtained from the frequent Establishment Censuses. Over time, a number of methods have been tested and used to keep the frame up-to-date in the years between the censuses. This is not an easy task, and very asymmetrical. Each and every closure of an enterprise, when selected for a survey, is easily observed as the unit can no longer be found operating. However, the establishment of new units, in general at a slightly higher rate than the closures, is not easily documented. The lists of new establishments registered by the chamber of commerce, the tax

Figure 1: Main Indicator of ES 1994-2012

authorities and the municipalities have been tried as sources but none of the efforts led to proper adjustments to the sample frame yet.

As a result, the number of establishments in the frame is getting reduced from year to year after each census as the closed units are taken off and in effect no new units are added (See Fig. 1). Thus, if only taking the information from the list frame, the number of units is shrinking from each census year and then boosted by a large percentage when a new census takes place. This problem has been founded

and the ES information for the period 1998 to 2006 has therefore been re-calculated using the later establishment censuses 2007, and 2011 after establishment censuses 2012, to find the appropriate adjustment factors. This work resulted in much improved data, especially for the years preceding the censuses as the distortions were the largest in those years (2003, 2006 and 2011).

After the 2007 CE, such a decline is not visible, as the economic surveys have been adjusted each year with the developments in enterprises as measured in the quarterly, labour force survey (LFS). This means that each year's economic survey reflects the total number of enterprises in the different activities and the corresponding employment, gross output and value added. The underlying assumption here is that the reporting and non-reporting units in a strata have the same value added and gross output per establishment. There is no reason to assume a bias either way, although it assumes that the average size of an establishment does not change over time.

The 2012 CE results gave numbers of establishments and their employees that differ quite substantially from the estimates for the same year based on the LFS. It was also found that the ES 2007 did not, actually, tally with the 2007 CE to which it was aligned. Thus, the ES series for the years 2007 to 2012 needs to be adjusted so that they match the information in the CE's for 2007 and 2012. There was a set of three scenarios to be considered for aligning the ES results, which were described as:

1. 2012 CE change in number of companies by year of establishment, by stratum;
2. Regression analysis; and
3. Imputing values for individual companies in years they were not included in the sample to expand "coverage" in each of the years.

The second and third methods were found to be not practical and therefore abandoned; the first scenario was selected. The adjustment procedures using this method covered the years 2009 to 2011. The number of enterprises existing in each year as reported in the 2012 CE is calculated. The adjustment factors are then calculated as the ratio between the number of establishments in successive years. This procedure was done separately for each of the strata listed below.

For the calculation of adjustment factors, the CE's for 2007 and 2012 were stratified by: Region (three regions); ISIC 4-digit and Employment size (3 groups), and the developments within each stratum analyzed.

Thus, if the 2012 census indicated that there were 100 units existing in 2008 in a certain strata, 110 in 2009 and 121 in 2010, the change from 2008 to 2009 would be (110-100) /100 = 10% and similarly (121-110) /110 = 10% for the following year. This change would then be applied to the data in the Economic Survey database for the respective years 2009 and 2010 for the selected strata.

This procedure has a major flaw as the unknown number of establishments that were started in the various years but had closed again before the CE 2012 have been excluded. The information given in tables 2 and 3 above do not make estimates of these cases, but given the rather high attrition rates and the large proportion of establishments in 2012 that were established after 2007, this is undoubtedly a sizable number. In other words, the biased series of the original ES series would be replaced by a series with a similar bias, also of unknown size. Therefore, another approach is needed and the LFS information was studied to evaluate the potential of the said series.

4. **Labour Force Surveys (LFS)**

Theoretically, the LFS reflects all employment in the economy. Applying filters to obtain only the part of the LFS data that matches the ES, should therefore provide proper information on the developments in the enterprise sector to correct the ES data for the reduced coverage over time due to closure of enterprises.

Basic filtering procedures

In order to obtain LFS information that matches the coverage of the ES, part of the LFS data need to be excluded from the comparison. For that purpose, a multi-step filtering procedure was applied as given below.

The LFS data are filtered as follows: Only persons of 15 years or older are taken and working in the local market.

Next, the type of employment is selected and in terms of question PW23 of 2009 and later years (see table 4) those responding that they are own account workers and wage employees in private sector are taken, that is, government workers and those engaged by international agencies, including UNWRA, are filtered out. Finally, persons engaged in agriculture and financial services are excluded. For this purpose, the individual data have been recorded in the new field of Industry ES.

LFS estimates of employment in establishments, 2007 to 2012

The results of these calculations, reflecting the estimated employment in establishments by region for the years 2007 to 2012 show large fluctuations as can be seen in Table 2 below.

Table 2. Estimated number of workers in establishments

Region	2007	2008	2009	2010	2011
WB	210,459	206,092	246,078	298,505	327,529
GS	65,854	44,843	54,797	75,561	108,161
Total	276,313	250,935	300,875	374,066	435,691
%age change		-9.2%	19.9%	24.3%	16.5%

While the employment stagnated in Remaining West Bank (RWB) in 2008, it went down very much in the Gaza Strip (GS). This is, of course, the result of the economic and political conditions at the time. Thus, even though the recent changes are very large, they include both a recovery from recent political disturbances as well as growth.

For the period 2007 to 2012, the LFS data have been analyzed and a specific format has been created to tabulate this data in a way that the employment presented matches the definitions used in the ES and CE. The 2007 and 2012 LFS data were changed from annual average estimates to information that comes closest to the reference dates of the respective censuses. The 2007 census reference date for employment was 30th September and the averages of rounds 46 and 47 of the LFS, which cover the third and fourth quarter of 2007, were used to match the census date. The reference date for the 2012 establishment census was 31st August 2012. The mid-point for the third quarter of 2012, LFS round 66, is 15 August which is closest to the CE reference date.

It was found that the LFS data for construction was completely out of line with the employment information in the construction survey. Actually, the number of employers and self-employed in construction listed in the LFS was broad at the same level as the overall employment recorded in the construction survey. The latter is a full census of the eligible contractors and therefore properly represents the activity of formal enterprises. The output of informal construction enterprises is estimated from the annual Existing Building Survey. In other words, construction workers are fully covered in the economic analysis of the national accounts. For this reason, the ES information for construction has been used to replace the LFS estimates for this activity.

The newly introduced classification of inside and outside establishments clearly created problems for the transport activity. For the early years, employment levels recorded in the ES and LFS were comparable, but for later years the employment in LFS was 3 to 5 times higher than that in the ES.

A probable cause of this is that the respondent to the LFS might confuse the transport operator's work with a taxi center as employment rather than, in most cases, self-employed outside establishments as the driver often owns the vehicle and just uses to services of the centre to obtain rides. The LFS data for transport have therefore also been replaced by those reported in the ES. While those adjustments affect two out of the six activities, which might seem a major intervention, it may be noted that the two activities together represent less than five percent of employment in the ES in all the years.

The adjustment factors have been prepared at the detailed level of the six surveys, even though some had rather few observations in the LFS. The construction and transport activities were not properly represented in the LFS sample due to their small footprint in the economy and difficulties their separation between employment inside and outside establishments. Thus, when looking at the development of construction, the results show that the estimation procedure is not adequate for this

activity. The number of establishments is estimated properly, but the employment shows large differences between LFS and CE for 2012 as well as 2007.

It may be noted that the calculations in this note are not comparable with the work done at the office. In this note the total employment in the enterprise sector forms the basis while the estimates used to adjust the economic surveys only took into account the development in the number of employers and self-employed in the establishments. The sum of those two more or less gives the total number of enterprises. Only since 2009 this information is asked for in the LFS questionnaire. As mentioned earlier, ratio analysis was used to extrapolate the inside-establishment estimates back to 2007. Of course, maintaining the same ratios for long periods of time, to include the earlier years does not help and therefore similar calculations have not been performed on the LFS data from 1994 to 2007.

Imputations to align all LFS datasets

The level of detail provided in the various data sets was not identical, such as ISIC codes at 2-digit level for 2010 and 2011 and 4-digit level detail in the other years. Furthermore, there were coding differences, such as the use of ISIC rev.3 for the years up to 2011 and ISIC rev.4 for 2012. Furthermore, the LFS questionnaire was modified in 2009 to some types of workers inside and outside establishments. The changes and imputation methods are described below. The questionnaires prior to 2009 did not separate the employers and self-employed persons working inside or outside establishments. This was changed in 2009 when the first two categories of workers were both split into two sub-categories, namely those working inside and outside establishments. The same differentiation was not made for employees or unpaid family workers.

The purpose of the analysis of LFS data is, of course, to find appropriate adjustment factors for the ES between census data. Preferably, such correction factors should be close to the results of the following economic census so that changes to time series remain minimal. From the confrontation of the published ES results over the past years and the employment derived from the LFS, it appears that this assumption is too optimistic. While the employment data from the two sources for 2008 and 2009 are more or less the same, for other years, the discrepancies are very large. The large discrepancies could be due to errors in abstracting the data from the LFS, but exactly the same parameter values have been used for all five surveys that were analyzed.

Table 3: Employment information for all ES activities, 2007-2012

Year	No. of Est.	ES-Employment	LFS- Employment	Adjustment Factor
2007	92,589	239,157	276,313	1.1554
2008	97,242	258,021	250,935	0.9725
2009	102,483	292,609	300,875	1.0282
2010	103,767	295,176	374,066	1.2673
2011	116,939	314,926	435,691	1.3835
2012	121,264	379,703		

Suitability of methods for ES projections

The tabulations above conclude the analysis of the LFS as far as the adjustment of the ES for 2007 to 2012 was concerned. The secondary objective was to test whether the method would yield results suitable for prediction so that in future the ES series does not need to be revised after every new Census of Establishments. In the following paragraphs, a description of the tests carried out to find a suitable correction method, is given. In summary, it may be noted that none of the tests carried out resulted in consistent results at the overall and the regional levels nor for the separate activities. Therefore, no series of LFS data stood out as the most appropriate one. This also means that it is not possible to provide an adjustment method to the ES results, using the LFS, that will overcome the problems faced in the past.

The LFS data were re-tabulated in various ways. First, one grouping was made whereby the businesses (the sum of employers and self-employed) were aggregated separately from employees

(paid or unpaid). As indicated earlier, the LFS questionnaire was changed from 2009 onwards. Employers and self-employed workers (business owners) were separated into two groups, namely those operating establishments and those not operating establishments but kind-of-activity units like taxis, mobile vendors, etc. The same was not done for the employees and unpaid family workers so that it was not possible to get a consistent division between persons engaged in establishments and outside establishments. The idea behind this grouping was that employers/self-employed represented the number of enterprises while the total workforce represented as closely as possible the overall employment registered in the CE.

It may be noted that all data of both LFS (2007 to 2012) and the series of 5 CE's from 1994 to 2012 were aligned with respect to ISIC codes (everything remapped to ISIC rev.4) as well as activity grouping (everything assigned to the six groups used in the latest ES) and regional classification (West Bank and Gaza Strip). This allowed for many levels of comparisons to see if any indicator series provided consistent results. Hereby consistency was not defined as accurate as such, but more in terms of similar differences across the various levels of detail (activities, regions and overall) for the years 2007 and 2012.

The LFS data were filtered so that workers in Israel and the settlements, as well as government employees and those working for UNWRA and international organizations were excluded. The employment definition used included those reporting to be working, those that indicated they were helping (generally unpaid family workers) and those having a job but not working during the survey period for various reasons. A check on the number of hours worked showed that the overwhelming majority of workers, including those "helping" reported working weeks of 35 hours or more, that is, full time employment.

A second method was to calculate the total number of persons engaged in private business, by broad groups of (a) employers / self-employed, (b) employees and (c) unpaid family workers. This comparison was done only at the overall and regional levels, but not by activity. At that level, the comparison was extended to include the 2004 Census as well.

As mentioned, none of the results give consistent results. The overall number of persons engaged, using 2004 as basis over-estimated 2007 and 2012 by 2% and 3% only, but at regional level the results were quite different.

The LFS estimates for the West Bank were seven to ten percent lower than those of the CE while they were up to 20% higher than the CE in the Gaza Strip. Patterns like this occurred at all levels, only in other distributions. Thus, no clear method to adjust the ES estimates between successive censuses presented itself, either at regional level or by activity.

5. Conclusion:

In summary, the various calculations show that no selection of LFS data over the period 2007 to 2012 provides a satisfactory projection of the levels of employment and number of enterprises found in the 2012 establishment census. In theory, the methods used should give fair estimates and the tests described show no obvious weaknesses in the LFS, but in practice, the 2012 estimates, as shown above for industry, do not match by a substantial margin. It may therefore be expected that use of the LFS to correct the ES, although better than not adjusting the ES at all between censuses, will require further adjustments after a new census has been conducted.

As described above, the LFS does not provide the type of information that provides an adequate solution for the adjustment of the annual ES prior to a new establishment census as it apparently does not reflect the changes in the business sector. In the following paragraphs, two lines of action are summarized. First, it may be possible that the development of a business register at PCBS provides the necessary information. Second, the implementation of a targeted small-scale survey may lead to the proper overall information for the enterprise sector.

References

Palestinian Central Bureau of Statistics, 2016. Labour Force Survey, Annual Report. Ramallah, Palestine.

Palestinian Central Bureau of Statistics, 2016. Economic Surveys, Annual Report. Ramallah, Palestine.