Making Tax Data Count - Uses in Official Statistics and Microdata Research

Slyuzberg, Michael New Zealand Inland Revenue Department 55 Featherston Street Wellington 2198, New Zealand E-mail: <u>michael.slyuzberg@ird.govt.nz</u>

Kelly, Neil Statistics New Zealand Statistics House, The Boulevard, Harbour Quays Wellington 6140, New Zealand E-mail: <u>neil.kelly@stats.govt.nz</u>

Hill, Hamish Statistics New Zealand Statistics House, The Boulevard, Harbour Quays Wellington 6140, New Zealand E-mail: <u>hamish.hill@stats.govt.nz</u>

Official statistics form the basis for informed policy development and evaluation by central and local government, as well as underpinning decision making by businesses and the public. Data collected within the tax administrative system offers opportunities to complement existing sources of official statistics, to support wider uses of available data, or to minimise provider load within existing outputs.

Why Tax Data is Important?

The potential role of administrative data for use in the production of official statistics is recognised in Principle 5 of the United Nations Fundamental Principles of Official Statistics:

Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents¹.

In the New Zealand context, administrative data use offers the opportunity to add value to the production of official statistics in a number of ways.

Tax administrative data sources in New Zealand tend to be almost universal for businesses, that is, they operate as an effective census. Because of this the data can be used to produce new statistical outputs related to small groups of interest, that would normally lie below survey design level – for example fine level geographical region or other sub-populations such as detailed industrial breakdowns. Administrative data in this situation becomes an efficient alternative data source².

¹ <u>http://unstats.un.org/unsd/methods/statorg/FP-English.htm</u>

² Dennis Trewin, The Future Role of a National Statistical Office. Wellington, 2007, pp 13-14.

This can be done at comparatively low cost and with no additional provider burden, whereas the cost of designing new sample surveys to query small sub populations can be a significant barrier to new collections.

Administrative data can also be used to lower respondent load in statistical surveys or collections that are already in existence.

In New Zealand, as in many other countries, businesses face demands for information from multiple arms of government – in many cases for substantially the same information. Although statistical collections represent only a small proportion of total respondent load, Statistics New Zealand faces the same challenges as other government agencies of reducing compliance load wherever possible. Reflecting this, one of the New Zealand government's statistical priorities is to "reduce costs to businesses and households through increasing integration, and reducing duplication³"

The use of administrative data in place of direct survey, particularly taxation data, is extremely important in achieving this goal.

The near universal and consistent nature of the New Zealand tax system means that a set of certain financial and business demographic information is available for almost all businesses operating in New Zealand – both annually and sub-annually. Given that this comprehensive set of information already exists, there would be no additional burden to statistical respondents, and minimal (if any) additional burden to taxpayers, if tax records were used for statistical and analytical purposes.

Summarising the above, use of tax data within the Official Statistics System supports the following key objectives:

- Supporting better informed policy decisions across the government.
- Facilitating new outputs at minimal marginal cost
- Meeting (within the legislative boundaries) information needs of other government departments and the requirements of international organisations.
- Reducing customers' response burden.
- Improving customers' voluntary compliance behaviour by:
 - o increasing the level of trust between customers and the tax authority;
 - o proper positioning of the tax authority in the community;
 - leveraging customers' compliance behaviour by sending appropriate signals to the community.

Demand on Tax Statistics in New Zealand and Overseas

In New Zealand, as well as most OECD countries, there is increasing use of administrative data for statistical purposes. Tax records in particular represent an important source of data for the compilation of official statistics. The official statistics and analyses derived from tax records are, in turn, an important source of data for economic and social research.

Our research shows that in New Zealand there is a genuine demand for tax statistics from policy makers, public sector researchers and academics. Examples of tax statistics on demand include:

- detailed income statistics (grouped by different demographic, geographic and social parameters);
- tax statistics for start-up businesses;
- tax statistics for micro businesses;
- tax statistics for Maori businesses;
- tax liabilities against tax paid;

³ Statistics NZ Statement of Intent; 2009-2012.

• social responsibilities statistics;

• tax statistics grouped by local government authorities, etc.

Many of the above statistics are published and are available for researchers and the general public from the New Zealand Inland Revenue Department (IRD) website⁴.

Tax data is widely used by statistical authorities in many other countries. For example, the US Census Bureau listed ten areas of its work that draw directly on federal tax data⁵: Survey applications branch; Census applications branch; Inter-censal population estimates programme; Longitudinal employer-household dynamics; Periodic demographic surveys; Small area estimates; Survey of business owners; Census/Social Security Administration joint research; External research; Economic Directorate.

The extent of this use, alongside data from other administrative sources, is a good illustration of the potential value of tax records in a range of statistical applications.

The Office for National Statistics (ONS) in United Kingdom uses tax data in two different ways – as a data source for published statistics, and in compilation of the sampling frame for statistical surveys of businesses⁶. Sample data from income tax and corporation tax records are an important data source for the national accounts and regional economic accounts; receipts data for taxes and national insurance contributions are used in compiling the public sector finances data releases. The data passed to the ONS are statistics derived from individual tax records, rather than any details from the individual records themselves. Therefore data privacy is strictly protected.

In future the ONS would like to make more use of tax data for statistical purposes in order to cut back on business surveys. The motive is to reduce compliance costs for business, particularly smaller businesses. The greater use of tax records, as a data source is one of the keys to improve regional economic data. The ONS set out proposals for an Integrated Population Statistics System, which would "combine census, survey and administrative data to create a single comprehensive population statistics database", but the time horizon currently envisaged for development is fairly long.

Tax Data Used in New Zealand Official Statistics

IRD supplies data to Statistics NZ purely for statistical purposes. As with the ONS, Statistics NZ uses tax data in two key ways- as a source of statistical information for publication, and in maintenance of the Business Frame.

Tax data is used indirectly to validate survey response for larger firms in economic surveys. The data is also used as a replacement for direct survey for small businesses in both annual and sub annual economic surveys.

The Business Frame is a register of all economically significant businesses operating in New Zealand containing life cycle, industrial activity and other classifications, and is the selection frame for all of Statistics NZ's economic surveys⁷. Data supplied includes:

• Information on tax registrations (business customers) – this information is supplied at least monthly and is used to enable Statistics NZ to maintain some of their core statistical datasets (e.g. Business Frame, Linked Employer Employee Dataset, Prototype Longitudinal Business Database). It contains data such as: business date of birth, physical location, industrial activity, type of entity, cessation date, etc.

⁴ http://www.ird.govt.nz/aboutir/external-stats/

⁵ UK Statistics Commission. Report No. 37. Tax Records as a Statistical Resource: A Review by the Statistics Commission. November 2007

⁶ Ibid.

⁷ More detail on the Business Frame is given in the later section.

- Information on tax returns this data supports producing a wide set of Official Statistics. It includes data returned by business entities, including their income tax, GST, "pay as you earned" (PAYE) returns as well as some financial accounts data. While the bulk of the data is supplied monthly, there can be multiple data supply events within a month.
- Information on student loan repayments and incomes Statistics NZ is the custodian of an integrated dataset on student loan borrowers which is used for valuation and modelling purposes. IRD supplies information annually on individual student loan borrower repayments, incomes, overseas status and loan balances for all borrowers since the scheme's inception. This is the only data in the above list which does not relate to businesses.

The volume of data supplied by IRD to Statistics NZ is considerable. For example, in 2009 IRD supplied 88.6 percent of all Business Frame administrative data received by Statistics NZ. In addition, IRD supplies approximately 120 million records per annum to the Linked Employer Employee Dataset (LEED) project team – this represents virtually 100% of the LEED data⁸.

It is estimated that overall IRD supplies around 175 million records per annum to Statistics NZ (excluding the Student Loan records). All of the data supplied by IRD to Statistics NZ is provided free of charge.

Needless to say that using tax records for official statistics purposes presumes complete protection of taxpayer confidentiality as a fundamental principle underpinning any tax agency. This issue will be discussed in more details in the following section.

Protecting Secrecy and Confidentiality of Tax Data

The working relationship between Statistics NZ and IRD are based on New Zealand legislative acts including Statistics Act 1975, Privacy Act 1993 and Tax Administration Act 1994. These Acts in turn protect the security of information provided to Statistics NZ, secrecy of tax data and confidentiality of taxpayers.

Pursuant to section 81(4)(d) of the Tax Administration Act 1994 the Commissioner of Inland Revenue is permitted to supply Statistics NZ with access to tax data provided that Statistics NZ will meet the following strict requirements.

- First and foremost, tax data should be used only for statistical purposes. This rule has no
 exceptions. Secondly, tax data supplied to Statistics NZ continues to be "restricted
 information" and as such remains subject to tax secrecy under auspices of the Tax
 Administration Act, despite the fact that it may be in the physical custody of Statistics NZ.
 This restriction applies to all data sourced initially from IRD, regardless of whether it is held
 at an individual or aggregate level. Finally, Statistics NZ's access to data sets that contain
 any data supplied by IRD are provided subject to the following conditions:
 - Any person granted such an access will be required to sign a secrecy certificate which certifies that they have been shown, have read and have understood section 87 of the Tax Administration Act prior to accessing a database. The signed certificates are retained by Statistics NZ as a permanent record.
 - Data sets remain within the Statistics NZ secure environment and the users are not permitted to export data outside this environment in any form (electronic or "hard" copying, emailing, using external memory devices, etc.).
 - All Information to be released is checked for confidentiality by relevant Employees of Statistics NZ, must be provided with a legal note referring to the requirements of the Tax

⁸ More detail on LEED is provided in the later section.

Administration Act 1994, and must never be taxpayer specific or allow taxpayer disclosure.

• All created interim/final data sets (both within mainframe and on any external memory devices) are to be destroyed after the project is finished.

If there is any question over whether tax secrecy applies to any particular Information or data, including aggregate level information, IRD will be consulted and the Commissioner's view on the tax secrecy of the matter will be accepted as final between the parties.

Managing Collaboration

Assisted by a supportive legislative environment, Statistics NZ and IRD enjoy long-term, wellestablished business relationships which help to organise collaborative work in a most effective and efficient way.

In 2004 Statistics NZ and IRD signed their first Memorandum of Understanding (MoU) which was then renewed in 2008. The Memorandum formally acknowledges that, in forging a closer relationship in areas of common interest, each party will be better positioned to achieve its own objectives and outcomes, and to contribute to achievement of the Government's current and future goals.

Based on the requirements of the Statistics Act 1975, the Official Statistics System model defines the areas of common interest for both Statistics NZ and IRD. The most relevant areas of common interest are:

- Identification of the key Official Statistics of national importance (Tier 1 statistics).
- The effective inclusion and participation of IRD as a recognised stakeholder of the Official Statistics System.
- Application of a consistent set of standards, principles and protocols for Official Statistics.
- Participation in the Survey Notification System (SNS) process for all surveys on the Official Statistics portal.
- In relation to areas of common interest, IRD and Statistics NZ will, where practicable, work together to:
 - Identify policy linkages between the parties and develop sound and consistent policy advice.
 - As far as practicable, seek to standardise policy and procedures affecting areas of common interest, bearing in mind the different focus and unique skills of each party.
 - Maintain necessary data exchanges.

The MoU notes that using tax data Statistics NZ in consultations with IRD will develop aggregated statistics, or the potential to produce aggregated statistics, that are publishable as meeting the confidentiality requirements.

The MoU also established a Coordinating Committee for discussing and resolving operational issues and other mutual relationship problems. In particular, the Coordinating Committee is tasked to perform the following functions:

- coordinate the supply and delivery of tax data;
- discuss and coordinate data quality issues;
- agree in principle to changes in data supply, policy and systems changes at an operational level;
- coordinate joint reviews of the data security;
- establish and maintain communication and understanding between the two parties;
- resolve any disputes or difference between the parties.

Whilst the legislative environment and inter-agency MoU provides a framework for the relationship between Statistics NZ and IRD, they do not by themselves ensure a successful working relationship. The willingness of personnel in both organisations to work together constructively within the defined bounds has allowed for increasing use of tax data to be made within this environment of mutual cooperation. Regular, close contact has enabled a shared understanding of benefits to be built, with the result that the supply of data from IRD to Statistics NZ is viewed as desirable for the statistics system as a whole.

Statistics NZ Products Using Tax Data

Business Frame (BF) Maintenance

The Business Frame is a comprehensive list of businesses and other undertakings in New Zealand engaged in the production of goods and services.

The information recorded about each business includes its name, address, ownership links with other undertakings, the numbers of persons employed, and codes indicating industry, location, institutional sector, type of business and nature of overseas transactions.

The BF is maintained to provide a framework for selecting respondents for Statistics New Zealand's business surveys. It is used to prevent gaps and overlaps in survey coverage, to manage the despatch and collection of survey questionnaires, to control respondent burden and to ensure consistent classification of data collected in business surveys.

The BF uses a three level statistical unit model comprising the enterprise, the kind-of-activity unit and the geographic unit. The primary unit on the BF is the enterprise and, in general, it is the enterprise that corresponds to a tax reporting unit. The enterprise represents a legal entity, which may be a company, partnership, trust, estate, incorporated society, producer board, local or central government organisation, religious organisation, voluntary organisation or self-employed individual.

To be in scope for the BF the business must fulfil any of the following conditions: have greater than \$30,000 annual GST turnover (or greater than \$40,000 income reported in the IRD Accounts Information form (IR10) if the business is not registered for GST); have more than three paid employees; be affiliated with other businesses; or be operating in the agriculture or forestry industries.

The BF relies heavily on the IRD's Client Registration database, comprising all taxpayers, excluding individual wage and salary earners, and is updated monthly by detecting changes in the Client Registration database. In addition, new IRD registrations which meet any of the criteria listed above are added to the frame monthly.

Extensive use is made of administrative tax data in maintaining the BF. This includes:

- Identifying new businesses that display over a minimum level of tax activity (an economic significance threshold) for inclusion on the BF.
- Creating the core BF record for new businesses using tax registration records. This includes deriving statistical classifications such as industry code, institutional sector and geographical code from tax system information.
- Updating employment and turnover size measures for existing businesses.
- Ceasing businesses on the BF, that are ceased in the tax system or display tax activity below the BF economic significance threshold.

For large and medium size businesses information collected by Statistics NZ is used to validate and extend the BF information, in addition to the tax information. Typically for small businesses, administrative tax information is the main source of the BF records.

Linked Employer-Employee Dataset (LEED)

The LEED integrated dataset is an ongoing data collection requiring monthly updates from IRD. Income derived from salary and wages has PAYE (pay-as-you-earn) tax deducted by the employer. All employers must file an Employer Monthly Schedule (EMS) monthly to the IRD. Large employers file twice a month. Every month the employer lists each employee and the employee's details. The LEED data base uses the EMS data as a base and adds industry and region information from the BF. In addition, information from the EMS data is used to repair longitudinal links between employers, in both the LEED system and in the BF system. This data and the history of updates to the BF are used to create the Longitudinal Business Frame (LBF).

The longitudinal nature of the dataset means that its value is increasing over time. As each month's data is added to the integrated dataset, a longer time series is being created without imposing any response burden on businesses or individuals.

Statistics NZ links IRD data on:

- those employers who make payments to individuals for work performed, and the individuals receiving this income.
- the equivalent statistical units drawn from Statistics NZ's Business Frame.
- additional payers of income with tax deducted at source, such as social security benefits, and their recipients.
- self-employed individuals identified from IRD reference tables. The LEED project is currently used for:
- Ongoing production and refinement of the dataset and statistical outputs.
- Investigation into linking the dataset with additional datasets, provided that full data linking policies and principles are in place, to better meet statistical needs.

The integrated dataset is used for statistical purposes only. It provides previously unavailable statistical information on job and worker flows, movement between earnings and income deciles and multiple job holding etc, and is used for:

- Production of official quarterly and annual integrated statistics.
- Periodical analytical papers by LEED team members.
- Official research by approved Employees of Statistics NZ including researchers seconded into Statistics NZ for bona fide research and approved by IRD.
- Restricted and documented internal use for validating other Statistics NZ outputs or maintaining statistical frames where no other source is available, for example AES, National Accounts and regional GDP and maintaining the Business Frame.

Longitudinal Business Database (LBD)

The LBD is a longitudinal database of integrated business-related data. It is enterprise-based and contains data from year 2000 (some data is available for earlier years).

The range of data includes information on business demographics, financial data, employment, goods exports, government assistance, and management practices.

The backbone of the LBD is the Longitudinal Business Frame (LBF, see above). Administrative data from other government agencies, a considerable amount of which is from IRD, is linked to the LBF, along with a number of Statistics NZ sample surveys that measure business practices and performance.

Tax data linked to LBD includes:

- Business Activity Indicator (BAI) a monthly series based on the supply of administrative data from IRD. The main source of this data is the GST (Goods and Services Tax) 101 form. GST is a tax based on the sale of goods and services.
- Financial accounts (IR10) collects a general summary of information relating to the business and its operations (profit and loss statement and balance sheet). IRD supplies IR10 data to Statistics NZ where it is transformed and linked to LBD.
- Company tax returns (IR4) is compulsory for businesses that are registered as companies. It includes income, tax calculation, refunds and/or transfers, provisional tax, and disclosures. IR4 data is supplied to Statistics NZ by IRD and is then linked to LBD.
- Linked Employer-Employee Database (LEED) see LEED section above.

It is the combination of the comprehensive administrative data with survey data which makes the LBD a valuable piece of statistical infrastructure to support a growing demand for micro-data research.

Economic Surveys

Having a tax data based business register (see The Business Frame above) enables Statistics NZ outputs to use tax data as an alternative to direct surveying. This has been undertaken in key economic surveys such as the Annual Enterprise Survey, Retail Trade Survey and other sub-annual collections.

The Annual Enterprise Survey (AES) is a postal survey of nearly all economically significant businesses in New Zealand. It collects detailed financial performance and financial position information for use in the production of annual National Accounts.

The population for the AES 2010 was approximately 450,000 business units. The AES population is split into two parts that use different collection methods. Tax data is used for sole proprietorships and partnerships, and selected small businesses. These comprise about 220,000 businesses (mostly sole proprietors and partnerships). The tax data used is sourced from the IR10 which collects about thirty variables from the statements of financial performance and financial position, including most of the financial variables needed by primary users, such as the National Accounts. The remaining 230,000 businesses are covered by a sample survey. Thus this sample, of 18,500 units, is selected to represent about half the population.

The benefits of this strategy are:

- Tax data was only used for smaller units that make a minimal contribution to measured aggregates.
- The tax stratum was full coverage, so the assumption was made that the reduction in overall sample error would be greater than any increase in measurement error resulting from the use of tax data.
- When very small businesses are included in a sample survey, units end up with very large weights. These large weights are a problem, if units change in an atypical way. The full coverage tax stratum eliminated the problems caused by small units with large weights.
- Experience using tax data for small units produced a better understanding of the quality of the data.

The use of tax data (GST) in the sub-annual economic collections has been in place for about the past nine years. The use of GST data has reduced the postal sample for the sub-annual collection quite significantly. The methodology of a mix of GST data and survey data has proven to be an effective strategy. The Quarterly Manufacturing Survey (QMS), Wholesale Trade Survey (WTS), and Retail Trade Survey (RTS) all use the same broad strategy for integrating GST with survey data. The

populations are sourced from the Business Frame and divided into three stratas: full coverage, sample, and tax stratas based on the size and characteristics of the businesses (using GST and employment data). Estimates are produced by aggregating unit-record data from the stratas.

The tax strata consists of single kind-of-activity (QMS/WTS) and single geographic-unit (RTS) enterprises below an identified size cut-off, targeting small to medium sized enterprises. The tax strata sources data from the Business Activity Indicator (BAI) which is an enterprise-based dataset created from transformed GST data. Data is produced on a monthly frequency in the BAI.

Micro Data Research

The legislative controls over the use of tax data and statistical data allow for the use of this information by researchers in other government departments for micro-data research under controlled conditions in Statistics NZ's secure data laboratory. The wide range of information in the LBD in particular has led to the production of a range of research reports examining various relationships between practices and performance of New Zealand businesses. Tax data in the LBD forms a key part of the classification of businesses and also the determination of performance metrics. Examples of papers produced from micro data research in the LBD include:

- By the Ministry of Economic Development:
 - Productivity Spillovers from Foreign Direct Investment in New Zealand: Firm Level Evidence;
 - o The Need for Speed: Impacts of Internet Connectivity on Firm Productivity;
 - o Firm-level Patterns in Merchandise Trade;
 - o Firm Dynamics, Market Structure and Performance;
 - o Comparison of Quantitative and Qualitative Firm Performance Measures;
- By the Reserve Bank of New Zealand:
 - Exporting and Performance: The Impact of Destination Characteristics on Learning Effects;
 - o Entrepreneurship And Aggregate Merchandise Trade Growth In New Zealand;
 - The "Suite" Smell of Success: Complementary Personnel Practices and Firm Performance;
- By the Department of Labour: Job Mobility and Wage Dynamics;
- By the New Zealand Transport Agency: Agglomeration Elasticities in New Zealand. Other agencies which have funded or used the LBD for micro data research are: IRD (as a data user);

Ministry of Agriculture and Forestry; Ministry of Foreign Affairs and Trade; Ministry of Research, Science and Technology; Treasury.

Learnings and Challenges

The use of tax data either in conjunction with or as a substitute for survey information has now been in place for some years in New Zealand. With such use increasing, experience by both Statistics NZ and those using the data for micro-research has provided valuable learnings, which in turn present challenges for the future.

- While tax data is generally universal and comprehensive, it is first and foremost collected for taxation purposes. Its use for other purposes means that the data is not always in the ideal format and sometimes additional data treatments are required.
- Different collection units present a particular challenge. Tax reporting units, especially in larger corporate groups, are sometimes different from the enterprise structures required for production of statistical aggregates. This can require apportionment across industries etc.

- The needs for collection of specific variables can be different when required for tax or statistical purposes. Definitions or inclusions for variables which appear to be the same (e.g. salaries and wages) may not always be so, requiring a clear understanding and sometimes changes to existing series or models to ensure fit-for-purpose use. Statistical agencies are often required to follow international frameworks and definitions in order to allow cross-country comparisons on equivalent bases. Often tax definitions for equivalent items differ across countries, as each jurisdiction looks to adjust its tax system to their own needs.
- Treatment of non-response can also be different, and often requires imputation to produce national statistical aggregates.

Where to from here?

Following initial use in the statistical infrastructure and surveys (as discussed above), an 'economics statistical architecture' has been created to provide the strategic foundation upon which Statistics NZ will transform the collection and production of statistics over the medium-term. As part of transforming its delivery of statistics, Statistics NZ intends to make better use of tax data to produce its statistical outputs.

Data will be directly collected (eg by survey) only when necessary. If this is achieved in economic collections, levels of business compliance costs will be reduced, particularly for small to medium sized enterprises. The associated efficiency gains in operational processes have the potential to decrease the costs of statistical production and provide an opportunity for further enhancements in the production processes through the development of effective statistical validation, analysis and output systems.

This architecture aims to support a developing set of information needs, including:

- A broader range of economic statistics to support policy development and monitor policy implementation.
- More reuse of data, with single data sets being used for a variety of purposes.
- Sharing of data across Government agencies.
- Improved microdata analysis.
- Quicker delivery of statistical information.

Implementing this statistical architecture is a long-term exercise. A key feature of this strategy is that decisions about the use of tax data will be quality driven. The continued quality of core statistical outputs will be ensured by expanding the use on tax data in incremental steps. Tax data will be used first in industries and sectors for which the data is known to be robust. Dependence on tax data will be expanded into other parts of the population as data issues are resolved. Suitable quality measures for mixed source collections will need to be developed.

In order to address these challenges, and those discussed under the previous heading, Statistics NZ and IRD will need to continue working together closely in future. As data collections are developed, variables and coverage can be better aligned. Emerging technologies can also offer new opportunities to make data supply more seamless or timely through electronic collection and transfer or real-time feeds.

REFERENCES

Fundamental Principles of Official Statistics, United Nations, 2006.

Dennis Trewin, The Future Role of a National Statistical Office. Wellington, 2007.

UK Statistics Commission. Report No. 37. Tax Records as a Statistical Resource: A Review by the Statistics Commission. November 2007

ABSTRACT

This paper looks specifically at the use of tax data in official statistics, micro data research and policy evaluation in New Zealand. Assisted by an environment of close cross-government collaboration, Statistics New Zealand and the Inland Revenue Department work closely together in using tax data to create new, policy relevant sources of information and extract value from existing datasets. The paper will cover benefits to the official statistics system in New Zealand from the sharing of tax data. The legislative and collaborative environment which allows for this to happen will be described, as well as the systems and processes of data supply.