



STATISTICS DIRECTORATE

STD/STESEG(2004)11
For Official Use

OECD Short-Term Economic Statistics Expert Group

Using administrative data in short-term statistics: Sub annual industry surveys at Statistics New Zealand

28-30 June
Room C
Château de la Muette, starting at 9:30 a.m. on the first day

Prepared by Rochelle Barrow
Statistics New Zealand

Submitted to the Expert Group under item 5 of the draft agenda

English - Or. English

JT00165547

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

USING ADMINISTRATIVE DATA IN SHORT-TERM STATISTICS

Sub annual industry surveys at Statistics NZ

Prepared by Rochelle Barrow
Senior Economic Statistician, Enterprise Accounts,
Statistics New Zealand

1 INTRODUCTION

Statistics NZ has recently completed a redesign of the monthly retail trade survey (RTS). This redesign completes stage 1 of the Sub-annual Economic Statistics Redevelopment Programme outlined at the STESEG conference last year¹. One of the objectives of this programme was to exploit the opportunities available from administrative data.

Over the past three years the quarterly manufacturing survey (QMS), quarterly wholesale trade survey (WTS) and RTS have all had new survey methodologies, IT systems and business processes introduced. All three surveys now make extensive use of administrative data. This data is used to update population details, periodically stratify units and replace direct surveying of small businesses.

The availability of additional stratification variables and the replacement of direct surveying of small business resulted in an estimated 25 percent reduction in respondent load across all three surveys².

Currently the threshold for the contribution of modelled administrative data to total industry estimates is set at approximately 10-15 percent. This threshold reflects the new nature of this kind of modelling at Statistics NZ. As experience grows with respect to the behaviour of administrative data it is intended that the thresholds will be reviewed and possibly raised.

This paper discusses the type of administrative data used in the sub annual collections, how it is manipulated and how it is used. It also discusses how some of the weaknesses of administrative data are managed by the three surveys.

¹ STD/STESEG (2003) 10 – Statistics New Zealand Background Paper, Nairn MacGibbon

² Respondent load was reduced by 25 percent in the QMS, 35 percent in WTS and 25 percent in the RTS

2 TYPE OF ADMINISTRATIVE DATA USED

The administrative data used in the three surveys is sourced from New Zealand's Inland Revenue Department (IRD).

New businesses, with annual turnover greater than \$40,000, must register with the IRD for Goods and Services Tax (GST) and deregister when they cease operation. The IRD has adopted this regime since GST was introduced in 1985. This includes the requirement that New Zealand businesses regularly file GST returns that are used to assess their tax liability. Depending on their size, businesses can choose to file GST returns monthly, two monthly or six monthly. Large business groups can choose to file individual GST returns for each legal entity within the group or one combined return for the entire group. Every filing period businesses report sales and purchases. This includes capital transactions (e.g. sale and purchase of plant) which are assessable for GST.

Salary and wage earners are taxed at source by employers who are required to submit an Employer Monthly Schedule (EMS). The EMS contains details of the gross amount paid to each employee and the amount of tax deducted. An EMS is filed monthly regardless of business size.

New Zealand businesses must file GST returns by the end of the month following the reference month (e.g. March GST returns must be filed by 30 April) and EMS returns by the 20th day following the end of the reference month (e.g. March EMS returns must be filed by 20 April).

The IRD extracts information on new businesses and GST and EMS data for Statistics NZ on the 18th of the second month following the end of the reference month. In most cases processing is not yet complete. Data processed after the 18th of the month arrives with next month's data file.

3 MANIPULATION OF ADMINISTRATIVE DATA

Administrative data must be incorporated into the Statistics NZ statistical unit model before it can be used. Significant time and resource was spent developing and building a concordance between the administrative frame (IRD number) and the statistical unit model. This concordance matches the IRD data received to the dynamic Statistics NZ Business Frame (BF). The BF is a register of all economically significant New Zealand businesses. It is used as the sample frame for almost all Statistics NZ business surveys as it also holds all stratification and selection variables. BF updating from administrative data occurs approximately 7-8 weeks after the end of the reference period.

In order to use administrative data as a substitute for direct surveying the data requires further manipulation. This manipulation is performed within the Business Activity Indicator (BAI) system which is currently used to produce an experimental output, based almost entirely on GST data.

Group apportionment

The first manipulation is to allocate data for group filers to each of the individual enterprises on the BF. This process uses additional administrative data, if it is available, or employment data to apportion GST data.

Estimation and apportionment of non monthly data

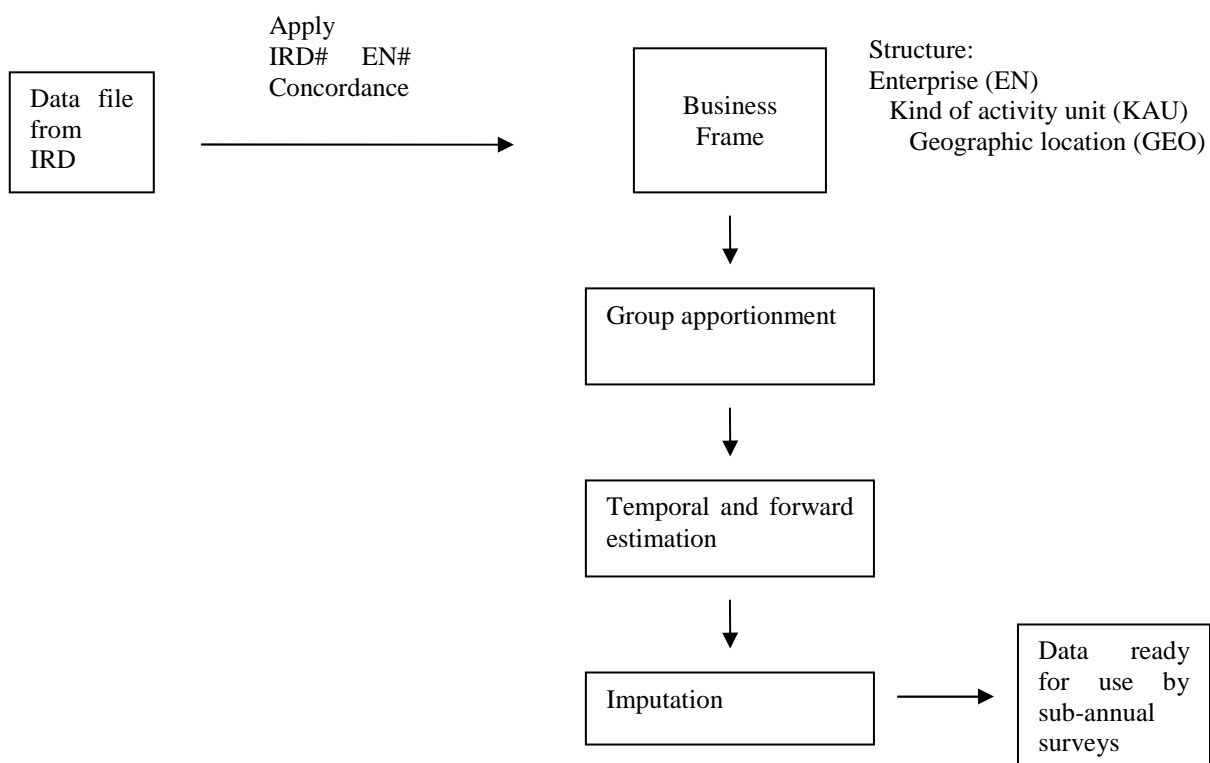
Businesses that file two or six monthly GST returns have their data estimated or apportioned across months in order to create a monthly series. In the months that two monthly or six monthly filing businesses are not expected to file the data is imputed. When data is eventually received it is apportioned across previous months using a series of industry specific temporal apportionment factors.

Estimation of missing data

The monthly data sets are at times incomplete. Processing delays and public holidays can impact on the amount of data available at the 18th of the month. Businesses that are expected to file, but are not included in the preliminary data set, are imputed.

The output of this process is an economy wide estimate of GST activity at the enterprise level. This series is an experimental output and the source of administrative data for the three sub annual surveys. Data is ready for use approximately 9 weeks after the end of the reference period. The diagram below summarises the process.

Figure 1



4 HOW ADMINISTRATIVE DATA IS USED

The use of administrative data in surveys requires compromise. The following section outlines how administrative data is used in the survey process. It also comments on how some of the weaknesses of administrative data are managed. These weaknesses include; lack of timeliness, statistical definitions that differ from administrative data source definitions, incomplete coverage of the variables of interest and missing data

Administrative data is used within the three surveys to:

- Update survey population details
- Stratify businesses
- Replace direct surveying of small business

4.1 Update population details

Administrative data supplied by the IRD is used to update the dynamic BF. New businesses, with annual turnover greater than \$40,000, must register with the IRD for Goods and Services Tax (GST) and deregister when they cease operation. This information is used to 'birth' and 'death' enterprises on the BF. Additional information including the industry of the business, its count of employees and location can also be updated from the administrative data source. However, business rules are used to determine which units can be updated automatically and which units need further investigation by staff.

The use of administrative data by the BF has reduced the resources required for maintenance and has improved the BF representation of the contemporary population.

4.2 Stratify businesses

The three surveys are stratified by industry for their sample selection. All businesses on the BF are industry coded (using ANZSIC - Australian and New Zealand Standard Industry Classification). For survey purposes ANZSIC classes are combined into industry groups called ANZIND³. Each ANZIND in each survey has its own bi-variate stratification boundaries.

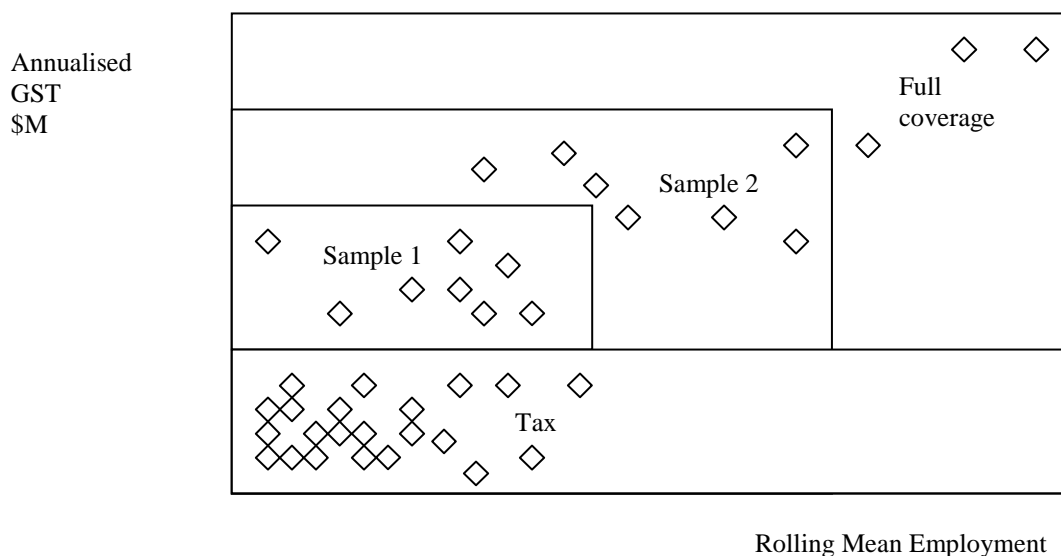
The stratification variables used are:

- annualised GST - the sum of the latest available returns that cover 12 months and
- rolling mean employment count - a 12 month rolling average of the count of employees recorded on the EMS

The use of two variables during stratification has improved the efficiency of the sample design. Below is a diagrammatic example of bi-variate stratification

³ There are 16 ANZIND in the WTS, 28 in the QMS and 24 in the RTS

Figure 2



All three surveys use periodic reselection. This means that every survey period all units in the population of interest are re-stratified based on the latest available stratification data. A new sample is selected and weights are recalculated and reassigned. In practice there is a large overlap of units between periods as the same stratification boundaries and selection rules are used.

The use of administrative data has resulted in a larger and more up to date set of stratification variables. Periodic reselection and regularly updated stratification variables enable businesses to 'grow' or 'shrink' and be representative of similar businesses, unlike panel samples. This ensures that an estimate at a point in time reflects the contemporary population.

4.3 Replace direct surveying of small business

One of the challenges facing Statistics NZ is the desire to reduce compliance costs, particularly of small and medium-sized businesses. With this in mind, the use of administrative data to replace direct surveying was introduced.

Each ANZIND has a tax stratum that contains small businesses that are not directly surveyed. Instead unit record administrative data is used for each business i.e. a census of small businesses.

Only single geographic location enterprises in the RTS or single kind of activity units in the QMS and WTS are eligible for inclusion in tax strata⁴. This ensures that further apportionment, below the enterprise level, is not required.

The use of administrative data as a substitute for direct surveying has four main issues:

- lack of timeliness
- the statistical definition differs from administrative data source definition

⁴ Multi geographic location enterprises can be included in the tax stratum if all geographic locations are in the same ANZIND.

- incomplete coverage of the variables of interest and
- missing data

4.3.1 Timeliness of administrative data

Tax data is ready for use by the surveys approximately 9 weeks after the end of the reference month. However, current publication time frames are shorter than 9 weeks. In order to achieve timeliness of the short term indicators administrative data is forecasted forward⁵.

For the quarterly collections actual data is available for the first two months and the third month of the quarter is forecasted. For the monthly survey data is forecasted two months ahead (e.g. data used in the May reference period is forecasted from data available up to March).

To address concerns about the ability of forecasting to reflect real world unanticipated shocks to the economy a modelling ratio is included in the RTS. GST sales are forecasted for all single geographic location businesses in the population of interest for the tax stratum and separately for the surveyed units. A modelling ratio is calculated using sales responses and forecasted GST sales for surveyed units and applied to the forecasted GST sales values for the tax units.

The ratio is ANZIND specific and only includes single location (RTS) or single activity unit (QMS and WTS) enterprises from the survey (to be more similar to those in the tax stratum). Any units deemed to significantly impact on the ratio are removed from the process.

$$\text{Sales for tax unit} = \frac{\text{total postal sales}}{\text{total forecasted GST sales for postal units}} \times \text{forecasted GST sales for tax unit}$$

4.3.2 Statistical definition differs from administrative data source definition

The administrative data source used does not conform exactly to the definitions of the variables of interest. The main reason is the inclusion of capital transactions in the GST sales and purchases variables. In order to adjust for this the data used in the surveys goes through a basic editing process within the BAI system and then more specific monitoring and editing of each surveys tax stratum. This process attempts to identify and remove potential capital transactions before the data is used in the surveys.

Due to the size of the businesses in the tax strata, and the fact that there is census coverage, any unidentified capital transactions are relatively small and unlikely to impact on estimates.

The modelling used in the RTS to adjust for unanticipated real world shocks also adjusts for any definitional differences that may occur between GST sales and sales reported in RTS questionnaires. The QMS and WTS do not use such modelling as it was not deemed necessary.

Statistics NZ has considered approaching the IRD with respect to changing the GST return form in order to separate capital transactions from sales and purchases. However, weighing up the potential

⁵ Data is forecasted using ARIMA models from the US Bureau of the Census X12 program. These models are calculated at an ANZIND level.

increased compliance cost and the marginal gain in the usefulness of the data we have decided not to pursue this course.

4.3.3 Incomplete set of survey variables

The administrative data received from the IRD does not contain all variables of interest, in particular stock levels. These variables are modelled using the administrative data and responses to postal surveys. An ANZIND specific ratio is used that only includes single location (RTS) or single kind of activity unit (QMS and WTS) enterprises from the survey. Any units deemed to significantly impact on the ratio are removed from the process.

The table below lists the administrative data used to model survey variables.

$$\text{stocks} = \frac{\text{total postal stocks}}{\text{total forecasted GST purchases for postal units}} \times \text{forecasted GST purchases for tax unit}$$

Table 1

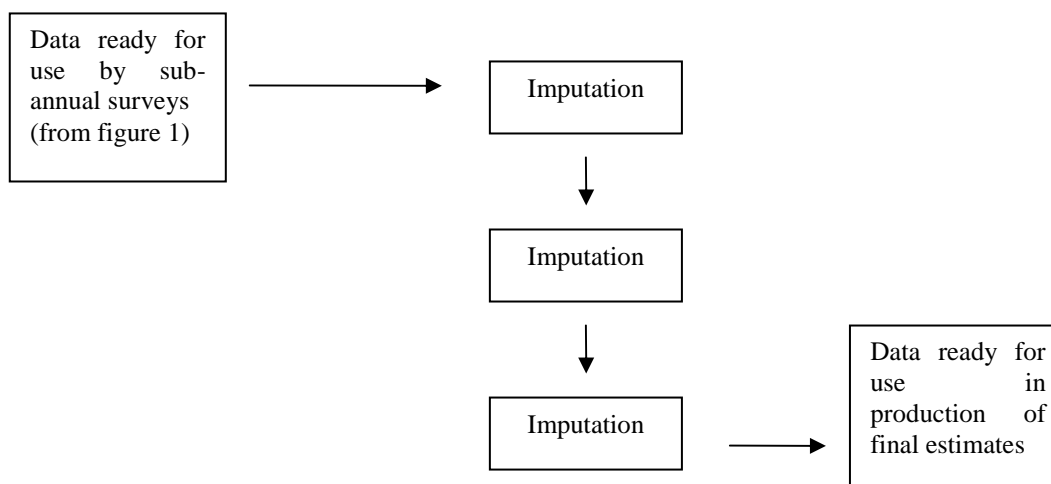
Survey variable	QMS	WTS	RTS
Sales	GST sales	GST sales	GST sales*ratio
Purchases	GST purchases	not in survey	not in survey
Salaries and wages	EMS salaries	not in survey	not in survey
Raw material stocks	GST purchases*ratio	GST purchases*ratio	not in survey
Finished goods stocks	GST purchases*ratio	GST purchases*ratio	GST purchases*ratio
Additions to assets	GST purchases*ratio	not in survey	not in survey
Disposals of assets	GST purchases*ratio	not in survey	not in survey

4.3.4 Missing data

The BF is used to determine the population of the tax stratum and therefore which units require administrative data. If GST or EMS data is not found for the required month it is imputed in a similar manner used for non response in a direct survey. This treatment can sometimes be required for new businesses. This method ensures that the population of interest for the three surveys has not changed from previous survey designs.

The following diagram follows on from figure 1 and summarises the above processes.

Figure 3



The table below summarises how potential weaknesses in administrative data have been managed in the three surveys.

Table 2

Weakness	Method	QMS	WTS	RTS
Timeliness	Forecasting	Third month of reference quarter - one month ahead	Third month of reference quarter - one month ahead	Reference month - two months ahead
Definitions differ	Modelling	Not deemed necessary	Not deemed necessary	Result of adjusting for unanticipated shocks
Incomplete set of variables	Modelling	Raw material stocks Finished goods stocks Additions to assets Disposals of assets	Raw material stocks Finished goods stocks	Finished goods stocks
Missing data	Imputation	Yes	Yes	Yes

5. CONCLUSION

All three sub annual industry surveys at Statistics NZ make extensive use of administrative data. The new methodology introduced into the surveys has increased the complexity of the surveys; however, it has also significantly decreased the respondent load placed on small businesses.

Weaknesses in administrative data have been successfully managed within the surveys by using forecasting, modelling and imputation.

6 FUTURE DIRECTIONS

Statistics NZ intends to review the thresholds for contribution of modelled administrative data to total industry estimates. We continue to monitor the behaviour of these new survey processes. As experience grows with respect to the behaviour of administrative data, forecasting and modelling the thresholds will be reviewed and possibly raised.

Statistics NZ continues to investigate the possibility of an Electronic Funds Transfer Point of Sale (EFTPOS) based indicator series. We are currently confirming requirements with data suppliers and hope to present some initial analysis at the STESEG 2004 conference.

7 APPENDIX

Information papers related to survey redesigns

- Changes to the Quarterly Economic Survey of Manufacturing
<http://www.stats.govt.nz/domino/external/pasfull/pasfull.nsf/7cf46ae26dcb6800cc256a62000a2248/4c2567ef00247c6acc256b260000ccb2?OpenDocument>
- Changes to the Quarterly Wholesale Trade Survey
<http://www.stats.govt.nz/domino/external/pasfull/pasfull.nsf/7cf46ae26dcb6800cc256a62000a2248/4c2567ef00247c6acc256cde007ac608?OpenDocument>
- Changes to the Monthly Retail Trade Survey
<http://www.stats.govt.nz/domino/external/pasfull/pasfull.nsf/7cf46ae26dcb6800cc256a62000a2248/4c2567ef00247c6acc256e2000751fe3?OpenDocument>
- Retail Trade Survey – Implementation of New Survey Design
<http://www.stats.govt.nz/domino/external/pasfull/pasfull.nsf/7cf46ae26dcb6800cc256a62000a2248/4c2567ef00247c6acc256e7000071b5e?OpenDocument>