Creating new and improved business statistics by maximising the use of administrative data

Chen Chen, Mathew Page, John Stewart[1]

Liability

The opinions, findings, recommendations, and conclusions expressed in this paper are those of the authors. They do not represent those of Statistics New Zealand, which takes no responsibility for any omissions or errors in the information in this paper.

[1] Chen, C, Statistical Analyst, Statistical Methods, Statistics New Zealand, Private Bag 4741, Christchurch, New Zealand (chen.chen@stats.govt.nz). Page, M, Senior Analyst, Statistical Infrastructure, Statistics New Zealand, Private Bag 4741, Christchurch, New Zealand (mathew.page@stats.govt.nz). Stewart, J, Senior Analyst, National Accounts, Statistics New Zealand, Private Bag 4741, Christchurch, New Zealand (john.stewart@stats.govt.nz).

Abstract

Statistics New Zealand has adopted an 'admin data first' approach in producing business statistical outputs. As well as reducing respondent burden, this new statistical paradigm allows us to dig deeper into the data which is a big win for our customers. We have developed an approach to maximise the use of goods and services tax (GST) administrative data to create efficiencies in our existing quarterly business statistics. GST is a comprehensive 'value added' tax collection undertaken by New Zealand's Inland Revenue. This approach has also allowed us to introduce a new range of turnover statistics for services industries, which has improved the quality of quarterly gross domestic product statistics. By expanding this thinking further and assessing additional administrative data sources, we are currently developing quarterly business profit statistics.

Keywords: Administrative data, services, profit statistics.

1 Introduction

1.1 This paper

Statistics New Zealand (Statistics NZ) is moving towards an 'administrative data first' paradigm in the production of economic statistics. The move was driven by the increased demand for business data and the pressure to reduce respondent burden and create statistical efficiencies. Statistics NZ's vision is to have administrative data as the prime source of business information, and to collect data (eg by survey) only when necessary[2].

The paper commences by introducing an approach we have taken to meet this 'administrative data first' paradigm with our sub-annual (quarterly) financial statistics. The approach delivered several benefits compared with the historic sample survey design.

Section 2 describes the application of an assessment model for GST sales data use within sub-annual business collections and the methods used to produce new sub-annual sales measures in manufacturing, wholesale trade, and selected services industries. GST is a comprehensive 'value added' tax collection undertaken by New Zealand's Inland Revenue Department that provides sales and purchases information.

Section 3 discusses some statistical challenges we have faced in making greater use of GST data in sub-annual outputs. These challenges include creating a quarterly managed collection for businesses where GST use is not suitable, developing methods for measuring quarterly inventories that are not in administrative datasets, and establishing ongoing statistical maintenance processes in the production environment.

Section 4 summarises our current work in expanding the new approach to cover more industries and variables in the future. Our aim is to produce quarterly business profit statistics across most industries in New Zealand's economy. The profit measures are being developed as part of the income-based measure of GDP that is being added to our quarterly suite of economic statistics.

The paper concludes with a short summary of what we have worked through in adapting the 'administrative data first' approach within sub-annual business collections, and what we will do next in furthering the approach to meet more customer needs.

---

[2] Statistics New Zealand Collection Strategy 2011, Statistics New Zealand, Wellington, 2011.

1.2 An 'administrative data first' approach for sub-annual business financial statistics

Statistics NZ has been using administrative data for a number of decades in the production of economic statistics. However, administrative data sources have been predominantly used to supplement survey data. In recent years we have seen fundamental changes to a number of our key economic outputs (that have relied on traditional sample survey design) and have move towards an 'administrative data first' paradigm. The development of an administrative data first approach for sub-annual business financial statistics is one of our recent achievements in making greater use of administrative data in economic statistics

Two main variables are currently required as part of the Statistics NZ sub-annual business collections – sales and inventories. These variables have historically been collected in three separate collections –Economic Survey of Manufacturing (ESM), Wholesale Trade Survey (WTS), and Retail Trade Survey (RTS). Statistics NZ publishes the sales and inventories information directly from the sub-annual business collections, but the data also flows through to the National Accounts and is used in the production of quarterly gross domestic product (GDP).

Before 2001 Statistics NZ's sub-annual business collections used a sample survey design. In 2001, a goods and services tax (GST) administrative data component was added to the design for the first time as a supplement to the sales survey data. The GST component was constrained to contribute no more than 15 percent of sales value for manufacturing and wholesale trade industries, and 10 percent for retail trade.

In the last two years, as part of Statistics NZ's business data transformation we have developed an approach to maximise the use of GST data in the production of sub-annual business financial statistics. This has resulted a new design for sub-annual business collections (see figure 1), where GST data (direct or modified by modelling) is used wherever possible, supplemented by a managed collection of large and complex businesses where the use of GST data is not suitable. For businesses in the managed collection we will continue to collect sales, inventories, and other key economic variables on a quarterly basis.

Figure 1 – History of sub-annual financial statistical designs
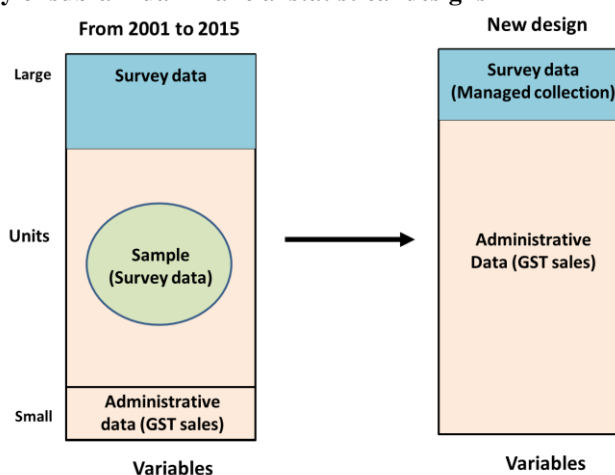
1.3 Key benefits of the new approach

The new approach is a significant paradigm shift in the collection and production of sub-annual business financial statistics, delivering the following benefits.

- **Reduction in respondent burden.** Maximising the GST data use in sub-annual outputs results in a significant reduction in respondent burden. We have achieved 50 percent reduction in respondent load across the combined sub-annual outputs.

- **Improvement in statistical quality.** The historic design for sub-annual business collection has limitations associated with a stratified random sampling. We found evidence that the sales series produced by the new design are of better quality, because it benefited from the GST data's near 'census' coverage.

- **Flexibility in statistical production.** The 'census-like' coverage of the GST data enables the production of estimates at lower levels of detail for the likes of ad-hoc research and in response to new customer demands. This was illustrated in the production of experimental sales series of Christchurch city retail trade and hospitality activity, following the Canterbury earthquakes of 2010 and 2011[3];

- **Scalability in statistical production.** We now have methods that enable the production of sub-annual financial statistics for industries not currently covered by Statistics NZ, such as services industries. The new outputs will improve the quality of our gross domestic product statistics.

- **Production of unit-record data**. There is an increasing need for microdata, which will underpin future statistical analysis. The new methods support the production of unit-record data for sales measures and potentially for other variables. This makes it easier to meet the need for micro-data analysis and the integration of data from different sources.

1.4 Implementing the new approach

The new approach has been implemented in the production of existing quarterly manufacturing and wholesale trade outputs. It has also been used to create selected services indicators[4], which were previously not covered by our sub-annual outputs.

The new statistical design and methods have been migrated onto our statistical production system, so that future designs supporting the new approach for a wider range of industries and variables can be implemented efficiently. Statistics NZ has moved to a single system for the processing and dissemination of a large number of its economic statistics. A key feature is the inclusion of survey and administrative data in the same system, along with the Business Register[5] which acts as a 'spine' to integrate many of our economic outputs.

By expanding the new approach further and leveraging off the new methods and system, we are currently developing quarterly business profit statistics across most industries in New Zealand's economy. This will significantly extend the existing industry coverage and variables to support measures of business profit and the development of quarterly income measures of GDP.

2 Maximising administrative data use in sub-annual business financial statistics

The approach taken has been to apply a model for assessing where and how the GST administrative data can be used as the primary data source in the production of sub-annual financial outputs. We have recently used the model to produce new sub-annual sales measures using the GST 'sales' variable for a number of industries.

---

[3] Christchurch retail trade indicator (information releases), Statistics New Zealand, Wellington, 2014

[4] The selected services indicators are within 'MN Professional, Scientific, Technical, Administrative and Support Services' and 'RS Arts, Recreation and Other Services' industries

[5] Statistics NZ's Business Register is a list of economically significant individual, private, and public sector businesses and organisations engaged in the production of goods and services in New Zealand.

2.1 An assessment model for GST sales data use

In ICES-IV we presented a GST sales assessment model (see figure 2)[6], which was used to assess the potential GST data use within our sub-annual business collections.

The key facets in the assessment model that enable decisions to be made included:

- the fitness for use between the administrative data and the statistical output – where the GST data can be used
- the methodological options available – how the data should be used.

The GST sales data is assessed by examining various aspects of the data such as reporting structures, timeliness, conceptual alignment, and reporting frequency. These are shown as decision boxes. Depending on the data assessment and its fitness for use, the most appropriate methodology is established for the various 'parts' of the business population. The data assessment and methods available underpin our aim of maximising the use of GST data, such that using GST data directly is the most favourable option and other data sources (eg survey) should only be considered when the use of GST data is not suitable.

Figure 2 – GST sales assessment model
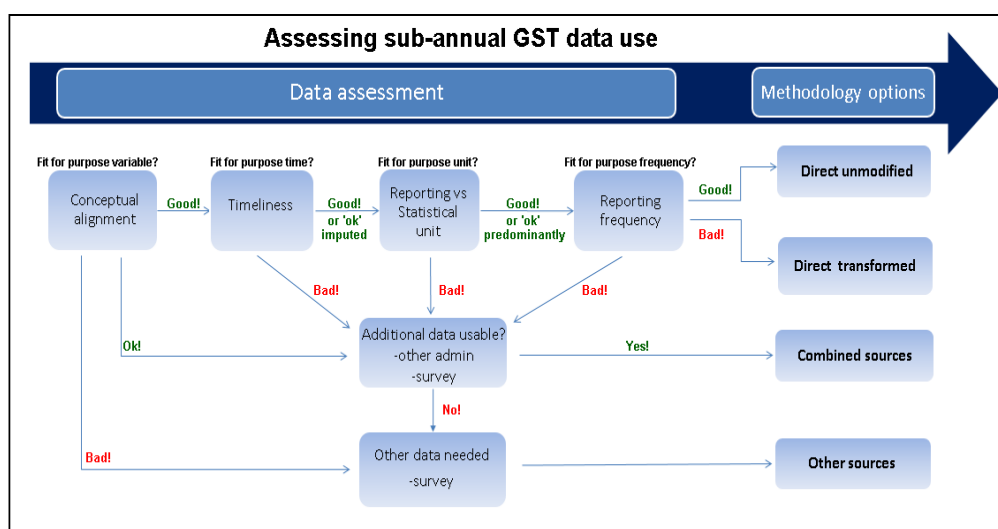
2.2 Applying the assessment model to produce new sub-annual sales measures

Statistics NZ has recently used the assessment model to produce new sub-annual sales measures for the existing manufacturing and wholesale trade industries. We have also released a range of new measures for selected services industries.

Table 1 presents the distribution of sales value and counts of businesses subject to each of the methods described in the assessment model.

Table 1 - Application of the assessment model by industry [(1)(2)] and method [(3)]

Method	Manufacturing		Wholesale trade		Selected services	
	Percentages of sales value	Count of businesses	Percentages of sales value	Count of businesses	Percentages of sales value	Count of businesses
Direct unmodified	12.1	1,370 (6.7)	18.7	1,630 (10.4)	11.2	1,310 (3.4)
Direct transformed (2 monthlies)	20.3	16,030 (78.8)	26.0	11,670 (74.4)	67.0	28,680 (74.3)
Direct transformed (6 monthlies)	0.4	2,520 (12.4)	0.6	2,050 (13.1)	5.0	8,330 (21.6)
Combined sources	0.9	110 (0.5)	0.5	90 (0.6)	1.8	70 (0.2)
Other sources	66.3	310 (1.5)	54.2	240 (1.5)	15.0	200 (0.5)

(1) Data is for the December 2015 quarter.

---

[6] Maximising the Use of Administrative Data in Sub-Annual Business Collections, J Stewart, Montreal, 2012.

(2) The data excludes any imputed units and units in industry FF116 'Commission-based wholesaling' (special methods are used).
(3) While only four methods are presented in our assessment diagram, this table separates the 'direct transformed' method by 2 and 6 monthly filers.

The process used to move through the assessment model will be further discussed in this section, with an emphasis on the challenges and outcomes in implementing the new approach.

2.3 Data assessment for GST sales use in selected outputs

2.3.1 Conceptual alignment

The assessment of GST fitness for purpose starts with assessing how the GST 'sales' variable aligns to the conceptual and definitional requirements of our sub-annual outputs. As noted, even though the GST data has not been collected for statistical purposes, we found (unsurprisingly) that the majority of the GST data can be used in the production of statistical outputs. The GST administrative data collected via the tax system is sourced from business accounts and therefore conform to our needs.

To test this alignment we take the annualised GST sale values and compare it with sales from our annual financial data. Where the evaluation shows the GST 'sales' variable closely matches our Annual Enterprise Survey (AES) definitions, the GST data can be used in one of the 'direct' methodologies that have been developed (such as the printing industry in figure 3). When there is a disparity between annualised GST and AES sales, the GST data cannot be used and another methodology is required. For example, in the 'commission-based wholesaling' industry, businesses can record their sales on either a gross or net (commission) basis.

Figure 3 – Assessing the conceptual alignment
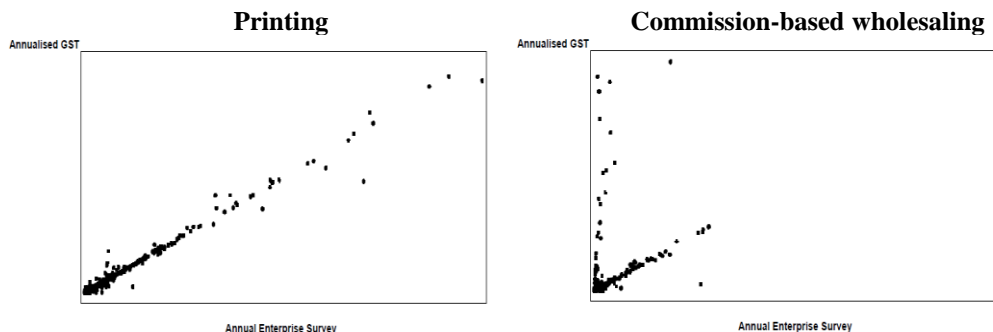
An additional conceptual alignment issue with the raw GST data is that it may include sales capital items, which does not align with the definitions needed by our users. The impact of capital items on levels and movements can be material and various from industry to industry. To solve this problem we developed a GST outlier strategy and specified rules to effectively remove large capital items from units in the manufacturing, wholesale trade, and selected services industries.

2.3.2 Timeliness

To use GST data in our sub-annual outputs we must have timely data. The administrative data must be available for the current period being measured. Use of administrative data solely from previous periods to model or forecast will not be fit for use in outputs that are needed to measure turning points and changes within the economy. Where GST data is only partly available for the current period, then fit for use data may be obtained through the standard imputation procedures. Where there is a significant proportion of the total value missing, then additional data may be needed to adjust the potential bias due to timeliness issues.

A significant finding in our assessment is the fact that at least 85 to 95 percent of GST data by value is available within current manufacturing and wholesale trade production timeframes. To ensure GST data can be used wherever possible, we developed a GST imputation strategy that reflects internationally recognised practice, based on Statistics Canada's Generalized Edit and Imputation System (Banff). Although the strategy has been preliminarily developed and implemented for manufacturing, wholesale trade, and selected services industries, it can be easily expanded to the wider economy.

2.3.3   Reporting structures

Having established the conceptual alignment between GST sales and the definitions in our sub-annual outputs and assessed the timeliness of the GST data, we then look at the suitability of GST reporting structure for our use. Administrative data is usually collected for the legal unit and this is the case with GST. Integrating the Statistics NZ Business Register with the legal units used for GST enables us to identify legal units that are most likely to be statistically fit for use. Such legal units are those that correspond to statistical units with activity in a single industry.

However, as the unit structures become more complex, the use of GST data becomes more problematic. For these units there is a decrease in quality and potentially an increased need for other sources to be used. An example is where Inland Revenue allow a group of legal units linked by ownership to provide data in a total that is collected against one unit while other units in the group record zero values (referred to as GST groups).

Table 1 shows that the vast majority of businesses by count (over 95 percent) have a simple reporting structure and the GST data can be used directly.

2.3.4 Reporting frequency

An issue that has to be addressed for using the GST data in our sub-annual outputs is the reporting filing frequency. GST data is reported to the IRD on several frequencies depending on turnover size. Higher turnover businesses report monthly, medium turnover businesses two-monthly, and low turnover businesses every six months.

For the quarterly manufacturing, wholesale trade, and services outputs considered here, a monthly reporting frequency is ideal. Data for the three months of the quarter can simply be added. However, businesses with a GST reporting frequency of more than a month require some form of transformation to enable the publication of quarterly data. This will be further explained in the methods section below.

2.4 Methods used for sales estimates in selected outputs

The assessment process for GST sales data use enabled us to determine the most appropriate method options for different parts of the business population.

2.4.1 Direct unmodified approach

The direct unmodified approach uses the GST sales data as it is received (with imputation and capital adjustment in place). This option is used when the administrative data reporting structures are suitable, the desired statistical outputs are closely aligned with the concept measured, and the reporting frequency enables timely use within the collection.

This is applied to GST monthly filers with a simple structure, and small GST monthly groups where the reporting unit or a single legal unit within the group accounted for greater than 80 percent of the group value. The sales measures are obtained by adding up the three-month GST value of these units.

The approach is a simple low-cost option and is fundamental to maximising the GST use in sub-annual outputs. Here the GST data is accepted in its own right as a measure of activity. Evaluating the quality of the estimates produced through this approach is straightforward, as actual data is used and therefore there is no uncertainty arising from the use of modelling and/or sampling.

2.4.2 Direct transformed approach

The direct transformed approach uses one part of the GST data that is itself fit for use, to estimate another part of the data where the data use is able to meet most our quality criteria but the reporting frequency is not fit for the purpose of producing quarterly statistical outputs. The GST data is used 'directly' in that it is not reliant on any other data source to produce the output.

This method is applied to the GST two-monthly and six-monthly filers with simple reporting structures, to deal with fact that neither the two-monthly data nor the six-monthly data can be added up directly to produce quarterly estimates.

- For two monthly filers we always have a month of GST data for the reference quarter missing, either the first month data or the third month data, depending on the reporting type of the filer. Modelling in conjunction with monthly filers is used transform the two-monthly data to quarterly data, with no other data sources involved.
- Six monthly filers can file GST returns in any two months of the year that are six months apart. That means we always have to deal with data filed in six different months and covers the reference quarter. The six-monthly data is apportioned and forwardcasted to the reference quarter when necessary.

By doing so we maximise our use of the available GST data while also avoiding any recourse to other data. Evaluating the quality of any aggregate estimates produced through this approach must be based on the model(s) used for producing such estimates.

- For the two-monthly data we have implemented a quality indicator in the selected outputs, measuring the variability in the aggregate estimates that occurs in the data transformation.
- The six-monthly data has been through several data manipulations, therefore evaluating the quality of the estimates is complicated. However, the contribution of six-monthly filers subject to this method is generally less than 1 percent in the manufacturing and wholesale trade industries and less than 5 percent in in the selected services industries – the estimates do not have material impact at the output industry level.

2.4.3 Combined sources approach

The combined sources approach uses the GST data in conjunction with data from other sources (administrative or directly collected), which are able to provide reliable benchmarks or useful auxiliary information to the GST data use.

The method is mainly applied to small and medium-sized GST groups where employment data from the Inland Revenue Employer Monthly Schedule (EMS), a monthly payroll return covering all employees, can be used to derive ratios to apportion GST sales data to the individual group members. This relies on the assumption that the employment data is an accurate reflection of the sales activity within the GST group. The quality of the apportionment process cannot be easily measured. To mitigate any quality concerns in a statistical production environment, we constrain the use of the method to less than 5 percent of total industry value. In

practice, the contribution has been less than this in the manufacturing, wholesale trade, and selected services industries.

2.4.4 Other sources approach

The other sources method uses non-administrative data sources to produce statistical output. This method is used when the GST data does not meet some basic quality criteria, and is therefore deemed as ineligible for our use.

For sub-annual business financial outputs, we found that using GST sales data for large and complex businesses is at times problematic and not fit-for-purpose. To ensure that we will produce feasible outputs for these business and maintain 'statistical control' over their reporting, we developed a strategy to create a quarterly managed collection for large and complex businesses where we will continue to collect their business information. The managed collection strategy is further described in section 3.

Although the count of units from other sources is low, they do make a significant contribution to total sales value, particularly in the manufacturing and wholesale trade industries.

3 Overcoming the statistical challenges of making greater use of administrative data

The major issues that needed to be resolved have been on businesses where GST use is not suitable and variables that are not in GST or other administrative datasets. Another issue at the production stage is how we maintain the quality of the estimates produced under the new approach.

3.1 Managed collection strategy

The managed collection strategy was developed to provide a framework to aid in the GST data use within sub-annual business collections. This framework identifies businesses where administrative data are not suitable for our statistical use. These businesses are included in a Statistics NZ quarterly managed collection where we will continue to collect sales, inventories, and other key economic variables that are required on a sub-annual basis.

A managed collection can be established based on the following principles

- **Significance** – The risks from using administrative data for businesses in a significant size include the inclusion of items outside the true conceptual measures for national accounting purposes, or any processing lags that do not meet our publication timeliness requirements. Statistical adjustments to deal with both of these facets for large businesses may have a significant material impact on our outputs.

- **Dominance** – A business needs might be important to an industry to the point where we would want to collect its business information to ensure the likely production of fit-for-purpose outputs.

- **Complexity** – Administrative data use for businesses with complex reporting structures or multiple industrial activities is often complicated and problematic.

As figure 4 illustrates, the key statistical issue for defining a managed collection for sub-annual business outputs is determining the boundary to maximise the use of GST data while continuing to maintain the quality of our outputs. We produced a number of simulated series using different business rules, and compared the results against historical published outputs. The business rules implemented in the production of manufacturing, wholesale trade, and selected services outputs are listed below.

- **A $100-million significance rule** – if an enterprise, or group of enterprises linked by ownership, have an annual GST turnover of more than $100 million.

- **A 3 percent industry dominance rule** – if an enterprise makes more than a 3 percent contribution to annual total income for an industry.

- **A structure complexity rule** -–all enterprises that have a significant level of activity across more than one industry.

Figure 4 – Managed collection strategy
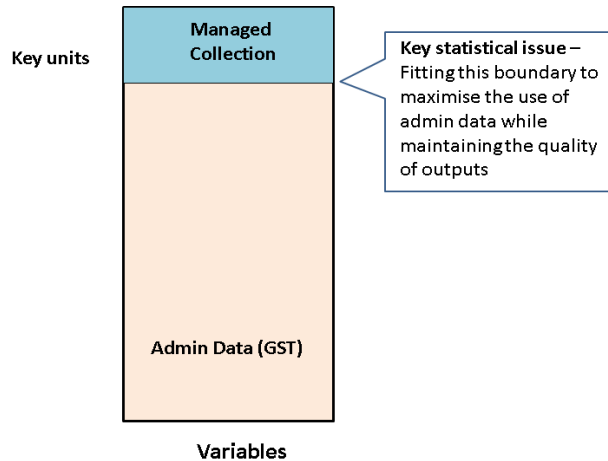
3.2 Measuring variables not in administrative data

A statistical challenge we had moving to the 'administrative data first' approach is how to produce statistical outputs for variables not in administrative datasets, such as inventories. Our approach is illustrated in figure 5. This framework identifies suitable methods to produce aggregate results for variables not available in administrative datasets, while continuing to maximise administrative data use. While this has been applied to produce inventories measures in existing sub-annual outputs, it could be used for other administrative variables across wider economic outputs.

Figure 5 – Sub annual business financial statistics – variable missing
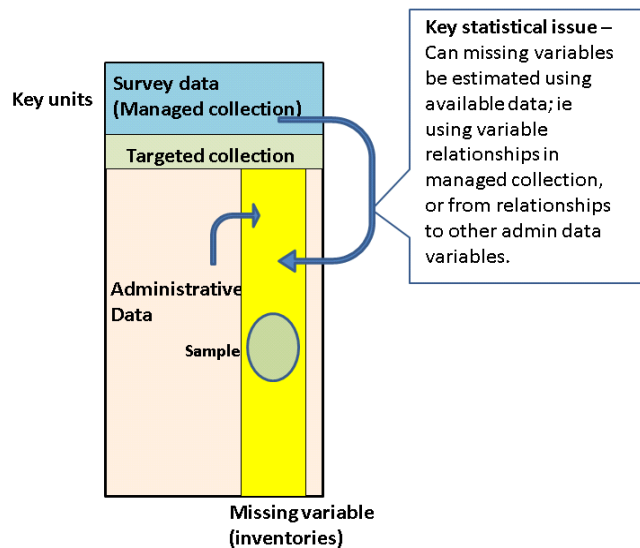
Quarterly inventory measures are required by national accounts in the production of GDP. We will continue to collect inventories data for businesses in the managed collection. We also conduct a small targeted collection to supplement the managed collection – we identify large inventory holders from our annual financial data that

are not in our managed collection and only collect their inventory information. However, there is no quarterly inventories data available for businesses not in the managed collection, as it is not collected in GST or other administrative sources.

Our analysis using historical sub-annual inventory outputs has found that different inventory estimation methods are required across the industries in our business population. When reviewing these methods, it is important to note that capturing the quarterly change in inventories was our primary objective (levels are a secondary consideration), and the complex and more resource-intensive approaches (eg sample survey) should be focused on the critical industries where inventory change has a material impact on economic measures. Like our approach for GST sales described in the previous section, we step through from our most favourable option to least desirable under an admin data first framework (eg sample survey)

- **Direct publish approach** – Inventory estimates are published directly using the managed collection data. This method is applicable where the managed collection businesses capture both quarterly change and levels in inventories.

- **Benchmark to annual approach** – Inventory estimates are obtained by 'rating up' the aggregate managed collection inventory series using annual financial data. This method is applicable where the managed collection businesses capture quarterly change in inventories and a significant proportion of inventory levels.

- **Model from managed collection approach** – Inventory estimates are obtained by modelling from the managed collection data. This method is applicable where the managed collection businesses capture quarterly change in inventories, but the levels obtained are not significant.

- **Model from annuals approach** – Inventory estimates are obtained by statistical models using the relationship between GST (sales and purchases), and inventories in our annual financial collection. This method is suitable for some smaller industries where the inventory levels remain relatively consistent over time.

- **Model within cluster approach** – Inventory estimates are obtained by modelling from the managed collection data using selected managed collection businesses in combination with GST data. This method is applicable where the managed collection businesses do not capture quarterly change and levels in inventories.

- **Sample survey approach** – If the above methods are not fit for purpose, a traditional stratified random sample survey is required. The sample only collects inventories data. This is applicable to larger industries where the managed collection inventory contribution is low.

When implementing the methods in the quarterly manufacturing outputs, we found only the 'benchmark to annual', 'model form annuals' and 'sample survey' methods were suitable for measuring of quarterly inventories. However, the other methods might be useful when we expand the new approach to a wider range of industries.

3.3 Establishing an administrative data quality framework

An assessment of the inherent quality of the administrative data and how well it aligns to the target economic variables or measures is imperative in an 'administrative data first' statistical design. There are two facets to an administrative data quality framework used by Statistics NZ:

- **Statistical design phase** – A data quality framework should be established as a key component of the initial design work. Our goal is to maximise the use of administrative data, and the framework is designed to identify which variables and which sections of the business population are amenable to measurement using administrative data. This work is completed in conjunction with the

development of statistical methods to transform the administrative data to make it fit for purpose. An example of this is the application of the GST assessment model described in section 2.

- **Monitoring phase** – Once the new design is in place it is important to have a quality assessment framework to examine the ongoing quality of the administrative data. The processes will be loosely based on the initial design framework, and should form a routine part of the statistical production process (eg benchmarking checks on an annual financial data). There should also be 'triggers' in place to re-examine any previously held assumptions – for example, changes in legislation which impact the administrative data source, or any event (eg a natural disaster) which may affect the timeliness of supply from businesses.

3.3.1 Ongoing statistical maintenance in the production of sub-annual business financial outputs

Under the 'administrative data first' paradigm, ongoing maintenance of the statistical designs and methods in the production environment will be vital to ensure the sub-annual financial series continue to reflect economic reality. A maintenance strategy we implemented in the quarterly production of manufacturing, wholesale trade, and selected services outputs is summarised below.

3.3.1.1 GST sales methods maintenance

The assessment process and methods are assessed annually using the latest sales information from our Annual Enterprise Survey to ensure that GST data continue to be suitable for our statistical use. The parameters and bias in the statistical models (eg transforming GST two-monthly and six-monthly values to a quarterly basis, apportioning group GST sales values to individual group members, etc) are monitored on a quarterly basis to ensure that we continue to produce sales estimates of a good quality.

3.3.1.2 Managed collection maintenance

We implemented an annual process to identify businesses to be forced into the managed collection. The process uses a ratio outlier procedure in Banff to identify any businesses where the annual sales from the latest AES and the GST sales during the same period are significantly different. This is followed by a correlation analysis looking at the effect each business has on the overall industry relationship between these two variables. The decision of forcing units into the managed collection is made based on the size of the detected units and how influential they are.

3.3.1.3 Inventories methods maintenance

For industries subject to the 'benchmark to annual' method, we need to confirm the managed collection continues to capture quarterly inventories change, and a reasonable and consistent proportion of inventories levels. This can be monitored by merging administrative data sources (such as GST and EMS) and managed collection data with the business population to ensure the managed collection contribution to other economic variables remains consistent. We can also use AES data (on an annual basis) to identify any potential large stock holders outside the managed collection.

For 'model from annuals' industries – the main issue here will be checking to ensure the relative size of the industries using this approach remains small. This can be assessed on a quarterly basis using GST, EMS, and managed collection data. We also need to ensure the integration of the updated annual parameters does not affect the derivation of inventories change statistics.

4 Expanding the methodological thinking in developing quarterly business profit statistics

Our next initiative is to significantly extend the existing industry coverage and content to support measures of business profit and the development of quarterly income measures. Statistics NZ is embarking on a significant upgrade to the scope and quality of our National Accounts, including the development of a quarterly income measure of GDP (GDPI) and balance sheets, to improve our range of economic statistics.

New data is needed to support these developments. We will need to extend our managed collection of large and complex businesses to additional industries in the economy. For the remaining businesses, we are investigating the use of GST sales and purchases in combination with salaries and wages from the EMS administrative data to derive profit estimates.

Two facets of the 'administrative data first' approach will be extended in this work. First, leveraging off existing methods and systems for maximum gain, and second, using this as a basis to develop a coherent multi-source collection including both administrative and manged collection data.

4.1 Leverage off existing methods and systems

The approach described in section 2 has been applied to our existing manufacturing, wholesale trade, and selected services industries to great effect to provide enhanced turnover and inventory statistics. We have moved from a sample survey collection of 1,800 businesses to managed collection of only 900 businesses. We have 78,000 businesses in our new design using administrative data only. The adoption and implementation of the new approach leaves us able to collect data at least cost from the maximum number of businesses, while also producing a wealth of business data.

The challenge we have is to expand this approach to derive maximum benefit for our customers in developing new business profit measures. We think the standardised approach already put in place can be efficiently scaled up to meet these new demands.

- Our Business Register is maintained primarily from administrative data and can be effectively used as the 'spine' to integrate additional data sources. The combination of administrative sources to calculate profit estimates will rely heavily on the statistical alignment of the data from both administrative data and managed collection sources. Statistical units from the Business Register and administrative units from GST and EMS can be effectively integrated as they are maintained from the same sources. We will experience minimal gaps and/or overlaps in coverage.
- The current managed collection strategy takes a standardised 'whole of economy' approach and can be easily extended to additional industries.
- The administrative data quality assessment frameworks (discussed in section 3.3) allows for scalability into further industries, and assessment of additional administrative data sources.
- Our innovative statistical production system has most of our economic data sources on the same system. This allows the standardisation of processes, methods, and tools across statistical outputs to increase flexibility and efficiency.

4.2 Developing a coherent multi-source collection to support profit statistics

We are working through a two-phased approach to developing new business profit statistics. We are currently working on enhancing the statistical design by expanding the administrative data quality assessment and managed collection process. Once the

methodological challenges have been resolved, we will integrate the additional sources and methods to our production system to support the new profit outputs.

4.2.1 Enhanced statistical design

4.2.1.1 Expanding administrative data quality assessment

To date our administrative data quality assessments have been confined to single economic variables. We have shown that GST sales is a good measure of 'sales' required to support sub-annual financial measures. We have also developed methods for producing a quarterly inventory variable which does not exist in administrative datasets. Our use of administrative data will now extend to multiple data sources being combined together. For our expanded range of outputs we need to use data on five key variables: sales, purchases, salaries and wages, inventories, and net profit.

Targeting an increased number of variables and deriving a complex variable such as net profit is a stern test of our administrative data first approach. GST purchases should be of similar quality to GST sales, although an assessment on the significance of capital purchases expenditure will be required. The integration of salaries and wages on the EMS should be relatively straightforward. The data assessment model can be expanded to include EMS-specific issues and potential method solutions. For example, the data can include lump sum payments that will affect our estimation (similar to GST capital spikes). Also, the salaries and wages data of a business could be reported on inconsistent pay periods (similar to GST filing frequency issues), so additional methods may be required to transform the data to a quarterly value.

There is no administrative data source that provides information on quarterly net profit. This was also the case for inventories and here we adopted a methodological approach to best estimate this variable. However, profit is a variable that is quite volatile and may not be readily correlated with our administrative data such as GST sales. Fortunately, in the case of net profit we can use the two administrative data sources together to proxy net profit. Thus, estimates of 'profit' will be derived from administrative data as GST sales less GST purchases less EMS salaries and wages. While the use of GST sales, purchases, and the EMS salaries and wages has the ability to provide a valid proxy for profit on a quarterly basis, further data assessment work will be required to ensure it is fit for purpose.

4.2.1.2 Extension of managed collection

The managed collection strategy is a standardised process that can be efficiently extended to other parts of the economy. This moves us away from data collection being largely output specific to a coherent, generic, integrated collection approach. However, this does mean that the managed collection has to be optimised to meet different variables or output needs. The existing managed collection rules are largely based on turnover measures that have been suitable for optimising the design for our existing outputs. However, with business profit being the key output variable under the extension, our generic process will require some modification in parts of the business population. Also, the data assessment process may identify some parts where the administrative data cannot be used to provide fit-for-purpose profit estimates. The challenge for us is to make suitable extensions to the managed collection where necessary, while continuing to maximise administrative data use. Following a similar process to our 'targeted inventory survey' we may survey businesses that make a large contribution to profit estimates but do not meet our turnover-based managed collection criteria. We could also consider further extending the managed collection, for industries that make a significant contribution to the profit measures and for which the quality of the administrative-based estimates are poor.

4.2.2 Integrating into the statistical production system

We have developed an innovative statistical production system that allows the integration of multiple economic data sources. Once the design and methods work has been further extended we will integrate the additional data, methods, and processes required to support the production of profit statistics onto the system. We already have processes in place to support the extended managed collection, and can broaden our existing GST processes to produce sale and purchases measures across the economy. We will require new methods to transform the EMS salaries and wages data to a quarterly frequency. The final step will be to develop processes to fully integrate the managed postal collection with profit measures calculated from the administrative data to produce profit statistics across the New Zealand economy.

5 Summary

Statistics NZ has undertaken the challenge to make greater use of administrative data in the production of economic statistics, to derive maximum benefit for our customers. We have recently implemented an 'administrative data first' approach to maximise the use of GST data in our existing sub-annual financial outputs. This approach has also allowed us to introduce a new range of turnover statistics for services industries which has improved the quality of quarterly GDP.

In this paper, we described how a GST sales assessment model was used to produce new sub-annual sales measures for manufacturing, wholesale trade, and selected services industries. We also discussed some methodological challenges we have faced in the process, such as managed collection strategy, measurement for quarterly inventories, and ongoing statistical maintenance.

Our next step is to further expand industry coverage and variables to support measures of business profit and the development of quarterly income measures in our economic statistics. We are currently working on enhancing our existing frameworks and methods by expanding the administrative data quality assessment and managed collection process.

References

[1] Chen, C and McGlone, S (2013). The Use of GST Data in Sub-annual Business Collections - Methodological Options. Statistics New Zealand, Christchurch (internal paper).

[2] Cope, J, Page, M, Stewart, J, & Chen, C (2016). Making greater use of administrative data in economic statistics. Asia-Pacific Economic Statistics Week Seminar Component, Bangkok.

[3] McKenzie, R (2008). Statistical architecture. Statistics New Zealand, Christchurch (internal paper).

[4] Page, M and Cox, N (2014). Furthering the use of administrative data in sub-annual financial statistics. International Association for Official Statistics Conference, Da Nang, Viet Nam.

[5] Page, M, Chen, C, & Li, C (2014). Future production of inventory measures in sub-annual financial statistics. Statistics New Zealand, Christchurch (internal paper).

[6] Stewart, J, Costa, V, Page, M, & Chen, C (2012). Maximising the use of administrative data in sub-annual business collections. International Conference on Establishment Surveys, Montreal.

[7] Stewart, J, Costa, V, Page, M, & Chen, C (2011). Towards an architecture for sub-annual business collections. Statistics New Zealand, Christchurch (internal paper).