Monthly Retail Trade Survey: Use of Administrative Data

The Monthly Retail Trade Survey (MRTS) is one of several business surveys conducted by Statistics Canada (STC) which generates estimates that measure the economic activity of Canada. This information is vital for governments and the private sector in their decision-making process. Respondent burden is an important issue faced by STC when conducting surveys. To mitigate some of this burden, STC has made great efforts in recent years to rationalize the data collected and orient surveys, where possible, towards the use of administrative data as a substitute for survey data.

The MRTS is one of the survey programs for which the use of existing administrative data (tax data) was tested and subsequently proven to be an adequate replacement of survey data. The objective of the Goods and Services Tax (GST) project was to use existing administrative data instead of survey data so as to reduce respondent burden and to lower collection costs.

As such, beginning with the October 2005 reference month, the MRTS will use both survey data and data modeled from GST returns to estimate retail sales levels. The following is a brief description of the methodology tested and eventually used in this process.

About the Goods and Services Tax and the Monthly Retail Trade Survey

Goods and Services Tax

The GST, introduced in 1991, is a federal tax levied on the consumption of goods and services in Canada. The tax is collected by the Canada Revenue Agency (CRA) for all provinces with the exception of Québec. All provinces, with the exception of Newfoundland & Labrador, Nova Scotia and New Brunswick, calculate the tax as a 7% charge on the value of the sale. In Newfoundland & Labrador, Nova Scotia and New Brunswick, the tax is a harmonized sales tax (HST) of 15%, which includes the GST and each province's sales tax.

All businesses, with the exception of those with revenues under $30,000, are required to file GST remittances. Businesses with annual sales revenue greater than $6 million per year must file monthly returns. Businesses with revenues between $500,000 and $6 million per year must remit quarterly. Businesses with revenues between $30,000 and $500,000 submit annual remittances. Monthly and quarterly reporters must remit within 30 days of the period end date, while annual reporters must remit within 3 months of their period end date.

The GST file is sent by CRA to Tax Data Division (TDD) at STC. TDD subsequently carries out further processing which is solely for statistical purposes at STC. This processing ensures a clean and complete database to be accessed by the various business survey programs at STC. The TDD processing includes the correction of erroneous data, outlier detection and replacement of missing data through calendarization and

extrapolation. The TDD processing is not intended to administer or monitor the GST program and no modifications are ever sent back to CRA.

Monthly Retail Trade Survey[1]

The MRTS is a sample survey which provides monthly information on sales representing retailers in Canada.[2]

In order to lessen response burden and to lower collection costs, the smallest units of the survey population (approximately the bottom 5% based on the dollar value of sales for each industry trade group by province) are excluded from being surveyed. This means that out of approximately 215,000 retail units in Canada, only about 100,000 have a possibility of being selected for the MRTS.

The MRTS sample is stratified based on industry, geographical region, and size (based on the annual dollar value of sales). The stratification regions consist of the provinces and territories, as well as three provincial sub-regions. Approximately 11,000 units are sampled from the 100,000 units in the sampling frame. The units remain the same from month to month, except for new units (births), which are sampled with the same probability as units in the original sampling frame.

Potential replacement of survey data by Goods and Services Tax data

The MRTS sample contains two types of units: simple and complex. For complex units (units consisting of more than one establishment), often only one GST value may cover all establishments. Consequently, the currently available GST data does not allow for breakdowns by geographical region and/or NAICS. Live complex units account for about 4,000 of the units in the MRTS sample and were thus deemed not eligible for replacement by GST data. The remaining units in the sample are simple units (units that represent only one establishment) and were deemed potentially eligible for replacement by GST data (about 7,000 live units).

The definition of simple unit applies to both small and large establishments. As a result, not all units that are deemed as simple are available for replacement. To ensure that the use of GST data does not have a considerable impact on the overall estimates, only units that were simple, live and not included in the units that represent the top 25% of an industry trade group or the top 25% of a particular geographical region in terms of sales are available for replacement. This ensures that simple large or dominant establishments will continue to be surveyed.

Linkage between Monthly Retail Trade Survey and tax database

---

[1] For more information on the methodology fo the MRTS, see **Definitions, data sources and methods: survey number 2406**

A link had to be established between the GST file and the MRTS sample in order to be able to replace the MRTS sales data with modeled tax data. The Business Number (BN) provided the link and a number of tests were conducted to see how many of the sample units were linked on a one-to-one basis to a GST record. This was the case for the majority of the units in the sample. However, there were situations where the GST revenue was zero for an establishment, while the reported data for sales was greater than zero, and vice versa. These particular cases were studied in detail to determine the source of the difference.

The October 2005 replacement units were chosen such that all of them were linked to a GST record.

Correlations between Monthly Retail Trade Survey sales and GST revenue

In order to determine whether the GST revenue data could be used as an adequate replacement for MRTS sales data, a good correlation between retail sales and GST revenue was required. The correlations when comparing the GST values of a particular month to the MRTS values of the same month are of good quality and even improve when the $0 values and outliers are removed.

However, due to timing constraints related to the release dates of the MRTS and the retrieval of tax data from CRA, the GST data are not available in time to be utilized by the MRTS for the current reference month. Data from the GST file that is one month prior to the MRTS reference month (e.g. February data for GST, March reference month for MRTS) is received in plenty of time for incorporation into the MRTS process. It was decided to verify whether a model could be designed using data from the GST file that was one month prior to the MRTS reference month. As illustrated in Table 1, the correlation between the current month's sales from MRTS and the GST revenue from one month ago was of good quality.

Table 1. Correlations between sales (March 2004) and revenue (February 2004)

Type of units	Correlation Coefficient
All units	0.5644
$0 reporters removed	0.9097
$0 reporters and outliers removed	0.9632

Simulation of the tax replacement process

Though the correlations between the MRTS sales and the GST revenue were very good, the primary concern regarding the use of GST data was that MRTS trends and current production levels would be preserved. To assure that the trends and levels were maintained, a simulation of the approach was performed for the period March 2004 to February 2005.

Identification of units to be replaced

The first step was to identify which establishments were available for replacement. These units are defined as simple, alive, and not in the top 25% of the domain of interest (industry trade group or geographical region). For the simulation, the domains consisted of industry trade groups (TGs) and geography (GEOs). The lone exception to this was TG 010 (New Car Dealers) which was excluded due to its large impact on the overall estimate.

As a result, 4,547 simple establishments were considered available for replacement for the first month of the simulation (March 2004). Simple units within other TGs and GEOs were excluded from being replaced since they were determined to be industry significant or they were units that were also surveyed by the Quarterly Retail Commodity Survey (QRCS). Units that are also surveyed by the QRCS (2,814 simple units in total) are not eligible to be replaced since it is difficult to determine commodity sales from the GST data. Due to units that are part of the QRCS sample and the units that were determined to be industry specific, a total of 1,636 simple units resulted as being the sample that would be replaced using a GST model. The remaining simple units are used to build the GST model. During each month of the simulation, only the remaining simple units that are deemed to not be outliers are used in the GST model that month to produce the modeled values for the 1,636 units being replaced.

Modeling

Modeling sales

For each month of the simulation, a model was created that was based on the sales values for the current month (M) and the GST revenue values of one month ago (M-1). The model was based on the sales to GST revenue relationship for units that were available for replacement but remained in the survey portion and were still alive. This model was then applied to the GST revenue values for those units that were selected for replacement.

To ensure that a proper model was created, outliers were removed from the model group. These outliers were observations where the relationship between GST revenue and the reported sales values was not like the majority of units in the same model group. These were cases where the GST value could be zero and the corresponding survey value for sales was a relatively large estimate, or vice-versa.

Once the model was generated, it was applied to the units to be replaced. Outliers generated by the model (units that are completely out of range when compared to all the units modeled) were removed and replaced with imputed data.

One of the main advantages of this approach was that once the values were modeled at the micro level, the existing estimation process could be used. As can be seen in Chart 1, the results of the simulation show little differences between the estimates containing administrative data and those obtained purely from the survey. The line that is denoted by 'Estimates based on survey only' corresponds to the estimates based on surveying all

units while the line that is denoted by 'Estimates using GST' corresponds to estimates based on surveying units other than those units that are going to be modeled based on GST data.   The 95% confidence interval (denoted by L-bound and U-bound) suggests that the 'Estimates using GST' estimates remain well within the bounds of the original estimate. This hypothesis can be applied to more detailed estimates, but it should be noted, with more industry detail, the modeling can be less dependable with outliers having more of an influence.

Chart 1 – Results of the simulation for total retail sales at the Canada level
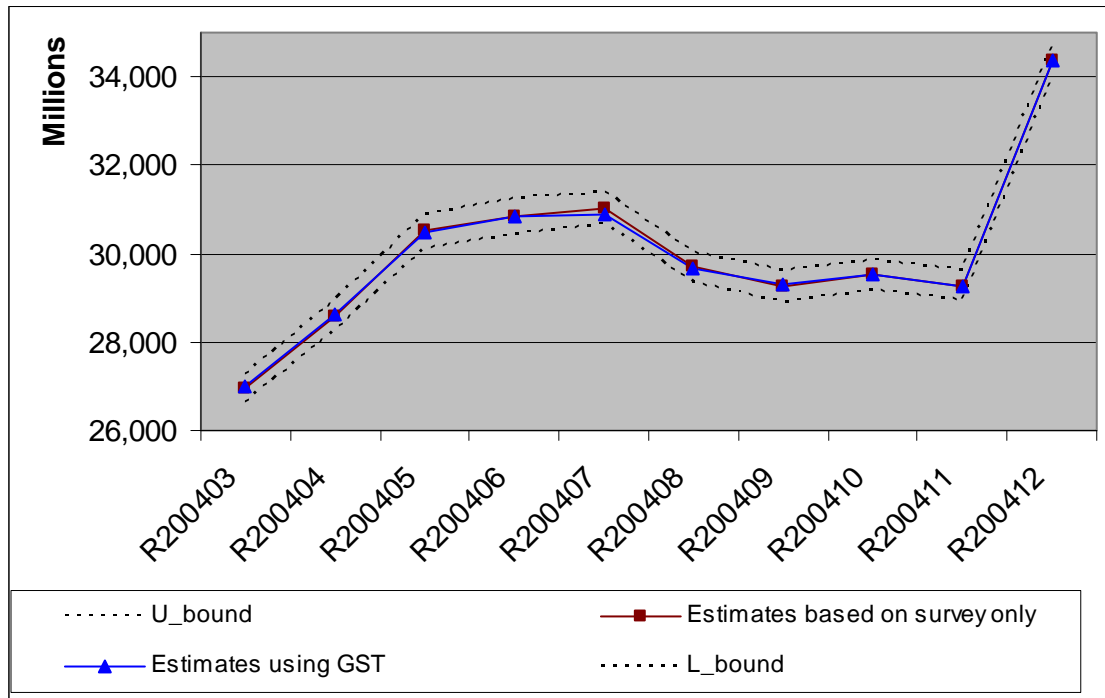
Conclusion

The replacement of survey data with administrative data, specifically GST data, was proven to conserve the high quality of the MRTS estimates while reducing respondent burden for small businesses in particular.   The new initiative also allows for more flexibility in selecting units to be surveyed every year as surveyed units can be replaced by administrative data without impacting the overall estimates.

As of October 2005, some 1,606 units (representing 14% of the MRTS sample) will not be sent a questionnaire. Rather, their values will be modeled using GST data for the sales variable. These results that have been obtained using the current model will be closely monitored in the future to determine if there are further opportunities for increased use of tax replacement.