Monthly Wholesale Trade Survey: Use of Administrative Data

The Monthly Wholesale Trade Survey (MWTS) is one of several business surveys conducted by Statistics Canada (StatCan) 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 StatCan when conducting surveys. To mitigate some of this burden, StatCan 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.

About the Goods and Services Tax and the Monthly Wholesale 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 Ontario, Prince Edward Island, Newfoundland & Labrador, Nova Scotia and New Brunswick, calculate the tax as a 5% charge on the value of the sale. In Prince Edward Island, 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. In Ontario, the HST is 13%.

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 $1.5 million and $6 million per year must remit quarterly. Businesses with revenues between $30,000 and $1.5 million 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 the Administrative Data Division (ADD) at StatCan. ADD subsequently carries out further processing which is solely for statistical purposes at StatCan. This processing ensures a clean and complete database to be accessed by the various business survey programs at StatCan. The ADD processing includes the correction of erroneous data, outlier detection and replacement of missing data through calendarization and extrapolation. The ADD processing is not intended to administer or monitor the GST program and no modifications are ever sent back to CRA.

Monthly Wholesale Trade Survey

The MWTS is a sample survey which provides monthly information on sales and inventories representing wholesalers in Canada.

In order to lessen response burden and to lower collection costs, the smallest units of the survey population (approximately the bottom 10% 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 100,000 establishments in the wholesale sector in Canada in-scope for the survey, only about 20,000 have a possibility of being selected for the MWTS.

The MWTS sample is stratified based on industry, province or territory, and size (based on the annual dollar value of sales). Approximately 12,000 establishments are sampled from the 20,000 establishments 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.

The MWTS has been making extensive use of the GST data for more than ten years now. The main reason behind the introduction of tax data was to reduce the response burden, especially for the mid-size businesses. Given that the MWTS contact their respondents on a monthly basis and that units remain in the sample for approximately five years, one can see that being selected in this survey is quite demanding.

In the early 2000's the MWTS was starting to see a decline in its response rates, especially in their mid-size businesses. The situation was becoming so pressing that something had to be done before those surveys were due for a restratification. It's important to mention that the MWTS goes through a restratification approximately every five years. In a nutshell, this process involves recalculating the strata boundaries and selecting a new sample. Of course, large units in the population are so important to the economy that they will always be part of the sample, even after a restratification. But the mid-size units are removed from the old sample and replaced in the new sample with randomly selected units from the population. Therefore, given that it was not time to restratify this survey, something else had to be done to relieve the mid-size businesses from their burden. The idea was put forward to use tax data to model their survey data.

Studies were conducted to see how well tax data was correlated to survey data for the mid-size businesses. Not surprisingly, the correlation between their income on tax data and their total sales reported on survey data was very high. Based on this result, the following strategy was adopted. It was decided that some of the smaller mid-size businesses who had reported stable sales since joining the survey would not receive a questionnaire anymore. These units were commonly referred to as the S2 units. Instead, their survey data would be imputed with a value obtained from a regular linear regression model between tax and survey data fitted to the businesses of similar size who did receive a questionnaire and responded to the survey. This second group of units was referred to as the S1 units. Overall, there were approximately 1,000 units in the S2 group. Although not methodologically perfect, this stop gap measure served its purpose and still provided very good quality data given the very strong relationship observed between tax and survey data. However, it was clear that this solution would have to be replaced eventually with a more sound methodological approach. In recent years, efforts were made to use tax data at the estimation step through the implementation of a ratio estimator.

Using a ratio estimator in monthly surveys is not a new idea. Studies on the topic have been conducted since 2000. At that time, GST data had been available for only a few years for use as auxiliary variables in surveys, and some concepts underlying these data were not completely understood or documented. In addition, the system for processing these data was not as well-developed as the one that is used now. Over the last 15 years, the methodology for processing GST data has continually improved, in terms of calendarization, imputation and allocation of business data to establishments. Everything is well documented and data quality is now excellent. Therefore, using GST data through a ratio estimator is now a promising avenue.

Ratio estimation consists of replacing the initial sampling weights (defined as the inverse of the probability of selection in the sample) by new weights in a manner that satisfies the constraints of calibration. Calibration ensures that the total of an auxiliary variable estimated using the sample must

equal the sum of the auxiliary variable over the entire population, and that the new sampling weights are as close as possible (using a specific distance measure) to the initial sampling weights.

For example, suppose that the known population total of the auxiliary variable is equal to 100 and based on a sample the estimated total is equal to 90, so that we are underestimating by approximately 10%. Since we know the population total of the auxiliary variable, it would be reasonable to increase the weights of the sampled units so that the estimate would be exactly equal to it. Now since the variable of interest is related to the auxiliary variable, it is not unreasonable to believe that the estimate of the sales based on the same sample and weights as the estimate of the auxiliary variable may also be an underestimation by approximately 10%. If this is in fact the case, then the adjusted weights could be used to produce an alternative estimator of the total sales. This alternate estimator is called the ratio estimator.

In essence, the ratio estimator tries to compensate for 'unlucky' samples and brings the estimate closer to the true total. The gain in variance will depend on the strength of the relationship between the variable of interest and the auxiliary data.

A nice feature of the ratio estimator is that it can be used to get an estimate for the whole population, including the smallest units excluded from being surveyed. This is done by simply including the take-none portion in the control totals for the sample portion. By doing this, the weights for the sampled portion will be increased in such a way that the estimates will be adjusted to take into account the take-none portion.

Correlations between Monthly Wholesale Trade Survey sales and GST revenue

In order to determine whether the GST revenue data could be used as an auxiliary variable for ratio estimator, a good correlation between wholesale sales and GST revenue was required. The correlations when comparing the GST values of a particular month to the MWTS 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 MWTS and the retrieval of tax data from CRA, the GST data are not available in time to be utilized by the MWTS for the current reference month. Data from the GST file that is one month prior to the MWTS reference month (e.g. February data for GST, March reference month for MWTS) is received in plenty of time for incorporation into the MWTS process. As illustrated in Table 1, the correlation between the current month's sales from MWTS and the GST revenue from one month ago is also of good quality.

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

Type of units	Correlation Coefficient
All units	0,8020
$0 reporters removed	0,8621
$0 reporters and outliers removed	0,9214

The use of tax data has gradually increased in the MWTS. What started out as very basic methods evolved through the years, culminating in the recent adoption of the ratio estimator.