Monthly Survey of Manufacturing: Use of Administrative Data

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Marketing and Dissemination Section, Manufacturing, Construction & Energy Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: (613) 951-9497, toll free: 1-866-873-8789, fax.: (613) 951-9499, e-mail: manufact@statcan.ca).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our Web site.

National inquiries line	**1 800 263-1136**
National telecommunications device for the hearing impaired	**1 800 363-7629**
Depository Services Program inquiries	**1 800 700-1033**
Fax line for Depository Services Program	**1 800 889-9734**
E-mail inquiries	**infostats@statcan.ca**
Web site	**www.statcan.ca**

Ordering and subscription information

This product, Catalogue no. 31-533-XIE, is available on Internet free. Users can obtain single issues at http://www.statcan.ca/cgi-bin/downpub/freepub.cgi.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136.

Statistics Canada
Manufacturing, Construction & Energy Division
Monthly Survey of Manufacturing Section

Monthly Survey of Manufacturing: Use of Administrative Data

Note of appreciation

Monthly Survey of Manufacturing: Use of Administrative Data

The Monthly Survey of Manufacturing (MSM) is one of several business surveys conducted by Statistics Canada (STC) which generates estimates measuring the economic activity of Canada. This information is vital for the decision-making process by governments and the private sector. 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 both rationalize the data collected and orient surveys towards the use of administrative data where possible as a substitute for survey data.

The MSM 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 August 2004 reference month, the MSM will use both survey data and data modeled from GST returns to estimate manufacturing shipment 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 Survey of Manufacturing

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 Statistics Canada. TDD then 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 Survey of Manufacturing[1]

The MSM is a sample survey which provides information on shipments, inventories, and orders representing all manufacturing establishments in Canada. Each month, the following characteristics are collected:
* Shipments (Goods of Own Manufacturing (GOM))
* Inventories (Raw Materials (RM), Goods-in-Process (GIP), Finished-Products (FP))
* Orders (Unfilled Orders (UO))

---

1. For more information on the methodology of the MSM see **Definitions, data sources and methods: survey number 2101**

In order to lessen response burden and lower the collection costs, the smallest units of the survey population, (the bottom 2% based on the dollar value of shipments for each province) are excluded from being surveyed. The strategy means that out of roughly 104,000 manufacturing establishments in Canada, only about 35,000 have a possibility of being selected to the MSM survey.

The MSM sample is stratified based on industry, province, and size (based on the dollar value of shipments). Approximately 10,000 units are sampled from the 35,000establishments 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 MSM 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, use of the GST does not allow for breakdowns by province and / or NAICS. Live complex units account for about 4,100 of the units in the MSM 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 5,008 live units).

The definition of a 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 80% of a particular NAICS industry in a particular province in terms of shipments are available for replacement. This ensures that large or dominant establishments, even those classified as simple, will continue to be surveyed.

Linkage between Monthly Survey of Manufacturing and tax database

A link had to be established between the GST file and the MSM sample in order to be able to replace the MSM shipment 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 sample. However, there were situations where the GST revenue was zero for an establishment, while the reported data for shipments was greater than zero, and vice-versa. These particular cases were studied in detail to determine the source of the difference.

The August 2004 replacement units were chosen such that all of them were linked to a GST record.

Correlations between Monthly Survey of Manufacturing shipments and GST revenue

In order to determine whether the GST revenue data could be used as an adequate replacement for MSM shipment data, a good correlation between manufacturing shipments and GST revenue was required. As illustrated in table 1, the correlations when comparing the GST values of a particular month to the MSM values of the same month are of good quality and even improve when the $0 values and outliers are removed.

Table 1. Correlations between shipments (July 2004) and revenue (July 2004)

Type of units	Correlation %
All units	0.826
$0 reporters removed	0.834
$0 reporters and outliers removed	0.984

However, due to timing constraints related to the release dates of the MSM and the retrieval of tax data from CRA, the GST data are not available in time to be utilized by the MSM for the current reference month. The previous month's GST data are also not received in time to allow for adequate review and implementation. Data from the GST file that is two months prior to the MSM reference month (e.g. May data for GST, for July reference month for MSM) are received in plenty of time for incorporation into the MSM process. It was decided to verify whether a model could be designed using data from the GST file that was 2 months prior to the MSM reference month. As illustrated in table 2, the correlation between the current month's shipments from MSM and the GST revenue from two months ago was of good quality.

Table 2. Correlations between shipments (July 2004) and revenue (May 2004)

Type of units	Correlation %
All units	0.779
$0 Removed	0.800
$0 and outliers removed	0.953

Simulation of the tax replacement process

Though the correlations between the MSM shipments and the GST revenue were very good, the primary concern regarding the use of GST data was that MSM 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: October 2002 to the most current month available at the time of the simulation.

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 80% of the domain of interest. For the simulation, the domain consisted of three-digit NAICS, by province. The exception was NAICS 336 (Transportation equipment) where the domain was four-digit NAICS, by province. Large or industry-significant simple units were also excluded from the list due to the impact on their specific industry or province.

As a result, 4,453 establishments were considered available for replacement for the first month of the simulation (October 2002). Since the impact of replacing these units was unknown and that some similar units needed to stay in sample to support the GST model, it was decided that 50% of these units would be eligible for GST replacement. Consequently, a simple random sample of half of these establishments was selected resulting in 2,241 units being selected for replacement.
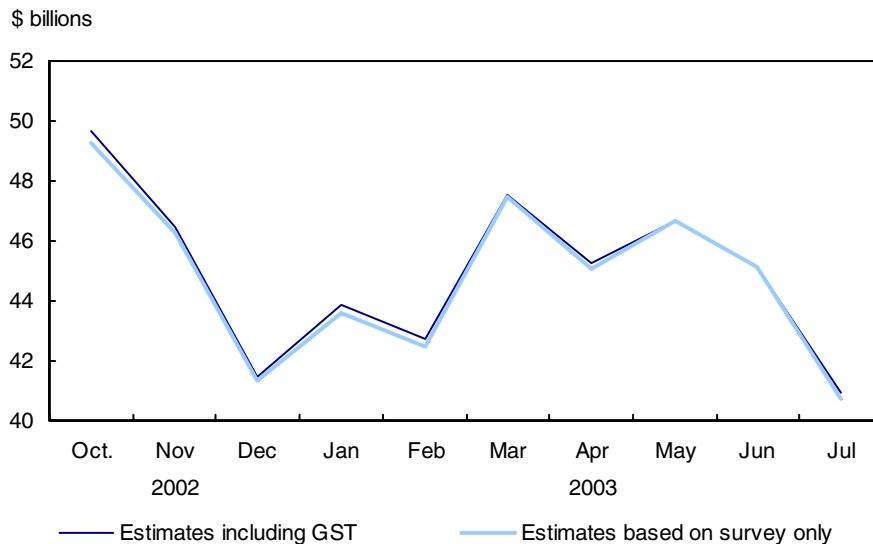
Modeling

Modeling shipments

For each month of the simulation, a model was created that was based on the shipment values for the current month (M) and the GST revenue values of 2 months ago (M-2). The model was based on the shipment to GST revenue relationship for units that were available for replacement but remained in the surveyed 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 GST revenue was not in-sync with the reported shipment values. These were cases where the GST value could be zero and the corresponding survey value for shipments was a sizable 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. With a 95% confidence interval, this suggests that the new values 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 is not as dependable and outliers can have more of an influence.

Results of the simulation for total shipments at the Canada level

Modeling inventories and unfilled orders

Currently, missing data for inventories and unfilled orders, due to lack of or late responses, is imputed by the MSM's processing system. The trend of establishments in the same NAICS / province group is used to impute for the inventories and / or unfilled orders of these units.

Various exploratory studies have shown little or no correlation between inventories and any of the variables available from the GST file. As a result, it was decided that the existing MSM

processing system will be used to impute inventories and unfilled orders of all establishments selected for GST replacement.

Although the quality of imputed data is adequate over a few months, with time, data quality of imputed inventories and unfilled orders is expected to deteriorate. Therefore, some precautions had to be implemented regarding long-term imputation. Since the MSM sample is re-stratified every three years, it was decided that three years would be maximum period that a GST unit would be kept before either being removed entirely from the sample, or rotated back into the survey population. For testing purposes, the inventories and unfilled order series were imputed for the 2000-2002 period and compared to the published series. The results of the test are illustrated in Chart 2. Similar results were obtained for the unfilled orders. The imputed data was of comparable quality to the series that was published, which led to the conclusion that imputation was an acceptable method of replacing survey data for inventories and unfilled orders for the GST units.

For units that become inactive (with a shipment value of zero), the imputation of inventories and unfilled orders will be stopped. Currently, when a survey-based establishment becomes inactive, depending on the size of the inventories, an analyst may decide to draw down the stocks of the unit over a period of a few months. Where GST units become inactive, the same strategy to draw down the inventories gradually may be implemented, if the size of the inventories is significant, although this is not expected to happen too often since most GST units are small. Therefore, some GST units may still have inventories for a few months after the unit has become inactive, otherwise, for most establishments imputation of inventories and unfilled orders will be terminated.

Results of the simulation for total inventory at the Canada level

$ billions

Simulated estimates   Estimates based on survey
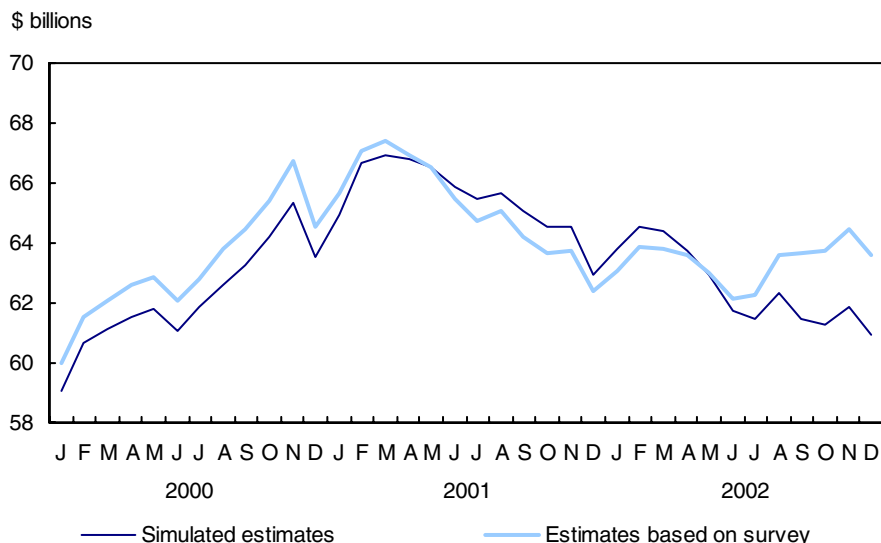
Conclusion

The replacement of survey data with administrative data, specifically GST data was proven to conserve the high quality of the MSM estimates while reducing respondent burden for small businesses in particular. This 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 August 2004, some 2,200 units (representing 25% of the MSM sample) will not be sent a questionnaire. Rather, their values will be modeled using GST data (shipments) and imputed data (inventories and unfilled orders). The results obtained based on the current model will be closely monitored in the future to determine if there are further opportunities for increased use of tax replacement.