

**Restratification and Methodology Changes in the
Monthly Restaurants, Caterers and Taverns
Survey (MRCTS)**

Implications and Impact on Data and Time Series

December 10, 2004

Table of Contents

Restratifcation and Methodology Changes in the MRCTS: Implications and Impact on Data and Time Series

	Page
Acknowledgments	3
Background	4
Reasons for Restratification	5
New Sampling Procedure	5
April 2004: Restratification / Resampling	6
Accumulated Frame Births	6
New Imputation System	7
Increased Use of Administrative Data with the Implemenation of the GST Model (Goods and Services Tax Data)	8
About the Goods and Services Tax	8
Potential Replacement of Survey Data by GST Data	8
Impacts of New Methodology on Estimates and Time Series	10
Revisions to Historical Time Series	11
Background	11
Revisions Process	11
Future Plans for the MRCTS	12

Acknowledgements

Statistics Canada invests considerable effort in collecting and compiling data for the Monthly Survey of Restaurants, Caterers and Taverns (MRCTS). As with any survey however, the cooperation of business operators in completing the questionnaire is the key factor to its success and, as such, Statistics Canada would first and foremost like to express its appreciation to the survey respondents.

A large number of people are involved both directly and indirectly in the production of the MRCTS. The following individuals from Service Industries Division are responsible for the processing of the data: Martha Degroot, Susie Knox, Alain Mbassegue, Joan Farnworth, Daphne Bennett. Significant collaboration is obtained from the Business Surveys Methods Division (methodology) and the Operations and Integrations Division (collection).

The following individuals from Statistics Canada are the principal authors of this Methodology report: Sean Crowe (Business Surveys Methods Division), Joan Farnworth (Service Industries Division), Jerry Fiori (Service Industries Division).

Restratification and Methodology Changes in the MRCTS:

Implications and Impact on Data and Time Series

Background

The MRCTS (Monthly Restaurants, Caterers and Taverns Survey) is a monthly survey that collects data and provides information on sales, number of locations, trends, and changes in the restaurants, caterers and taverns industry in Canada. These data are used by federal and provincial governments, industry associations, and food service businesses for consulting, marketing and planning purposes. The provincial and federal governments use the information to estimate provincial taxation shares.

The target population consists of all statistical establishments with employees (sometimes referred to as firms or units) classified as food services and drinking places according to the 1997 North American Industry Classification System (NAICS 722). A statistical establishment is a production entity which produces as homogenous a set of goods or services as possible, which does not cross provincial boundaries. There are approximately 65,000 establishments classified to NAICS 722 in the sampling frame. The frame is maintained by Statistics Canada's Business Register (BR).

The sampling unit for the MRCTS is the statistical company, which is an organizational unit above the establishment. The MRCTS sample is stratified (grouped) based on industry, province, and revenue size. Estimates are produced at the province/territory level for each of the five kinds of business (KOB). KOBs are a grouping of 4 and 5-digit NAICS codes, as tabled below:

KOB	NAICS	Description
1	722110	Full-service restaurants
2	722210	Limited-service restaurants
3	722310	Food service contractors
4	722320 & 722330	Caterers and mobile food services
5	722410	Drinking places (taverns)

In terms of business activity, the food services population can be volatile, where it is not uncommon for units to change their KOB classification over time. In addition, the demarcation between classifications is not always clear. For example, if a 'tavern' takes in a large proportion of its revenue from food sales, it might report these sales as a full-service restaurant, a tavern, or both.

Once the MRCTS sample is drawn, it remains the same from month to month. New units that have been birthed to the frame are sampled each month with the same probability as units in the original sampling frame. Dead units, both in-sample and out-of-sample, are retained as part of the sample control file (SCF).

A unit is assigned to a stratum in the month when it initially appears on the MRCTS sample universe file (SUF). Once a unit is assigned to a stratum, it will remain in this stratum until the time of the next

restratification. Similarly, the sample status of the unit (in sample or out) is determined at this time and cannot change until the next resampling.

In the MRCTS, there are two kinds of sampling units: Must-takes and take-somes. Must-take units are placed in a special take-all stratum whereby all such units are selected into the sample. The take-some units are placed in strata, where the strata are defined by province || KOB || ZONE. A ZONE is a size measure that is a function of a unit's province, KOB and annual SUF revenue. There are three ZONES in each province / KOB corresponding to three strata: a take-all stratum for large revenue units, and two take-some strata for medium and small revenue units.

At the time of restratification/resampling, all frame-dead units are removed from the SCF, and the restratification is performed using the most up-to-date information available.

The MRCTS received its last restratification/resampling in January 1999. The present restratification/resampling was implemented for the reference month April 2004.

Reasons for Restratification

The primary reason for the present restratification was simply the age of the stratified frame. The frame had inevitably degraded over the five years since the last restratification (a large proportion of units reporting in provinces/KOBs outside their sampled stratum). Indeed, most monthly surveys have a restratification every one to three years. As well, the MRCTS has used the same basic systems for sampling and estimation since the mid-1990s.

A subsidiary reason was the implementation of the Goods and Services Tax (GST) model into the MRCTS. This would allow some simple MRCTS sample units to have their sales values modelled instead of collected, saving collection costs as well as reducing response burden. It was determined that the GST model could not be properly implemented without a restratification.

New Sampling Procedure

There were several changes in the MRCTS sampling procedure:

- 1) The criteria for a unit to be placed in the must-take stratum have been strengthened. In both the old and new systems, all territory units (i.e., in the Yukon, Northwest Territories, and Nunavut) and all franchise collection units are automatically placed in the must-take stratum. However:
 - a. In the old system, the unit is automatically a must-take if it has multiple establishments on the SUF.
 - b. In the new system, a unit is automatically a must-take only if it has establishments in more than one province || KOB domain on the SUF.

This change was made to reduce overall sample size. A larger portion of the population will be sampled, rather than being the subject of a census.

- 2) The ZONE boundaries in each province || KOB domain were adjusted to take economic changes into account and optimize sample allocation.

- 3) The sampling fractions in all take-some strata were re-calculated in order to optimize sample allocation.

April 2004: Restratification / Resampling

The MRCTS was restratified for the April 2004 reference month. The April 2004 MRCTS SUF was updated through subject-matter research, including the removal of dead units. New size boundaries were calculated using the newly created Statistics Canada product, StatMx, in order to optimize sample allocation based on the desired coefficients of variation for the various estimates. The updated SUF was then stratified and sampling weights for each stratum were determined utilizing the Generalized Sampling System (GSAM). GSAM took into consideration the overall target sample size and the desired CVs.

A new sample was selected from the updated and stratified SUF. The new sample was chosen using GSAM in order to give 'maximum overlap' with the old sample. By definition, the new April 2004 SCF contained no frame-dead units.

An old version of the April 2004 sample was also selected according to the formula specified in the preceding section (taking the old March 2004 sample and adding a sample of Apr2004 frame births). Data were collected, and estimates were produced, for both the old and the new April 2004 samples in order to measure the restratification impact. The overlap of the old and new estimation series was one month only, April 2004.

Accumulated Frame Births

New businesses are added to the Business Register (BR) on a regular, ongoing basis. Occasionally, additional research is required to validate certain units before they are added to the BR. Depending on the nature of the research and the number of units involved, such units can accumulate. Accordingly, a backlog of new units (which should have been birthed in previous months) was added to the BR SUF in February 2004. These units should have been birthed onto the SUF in the months prior to 2004, instead of all in February 2004. Some units could actually have commenced operations prior to 2003. As a result, the implementation of these births were excluded from previously published MRCTS data until revisions to the historical time series could be made to reflect all the changes discussed in this paper.

New Imputation System

In conjunction with the restratification, the MRCTS sampling and estimation programs were reviewed, particularly the imputation system. Imputation is applied to records that have missing data. The imputation system prioritizes and automatically selects the appropriate method depending on the availability of the auxiliary data. Possible imputation methods are based on a record's own historical data, month-to-month trends, year-to-year trends, prorated annual data, and data from alternative administrative sources.

The MRCTS methodology and programs had remained relatively unaltered since January 1999. Some portions of the methodology, in particular the imputation system, were much older still. Thus, it was decided to implement a revised imputation procedure.

There were several changes made to the MRCTS imputation process:

- 1) The imputation is performed according to survey domain instead of sampling stratum. The survey domain(s) of a sample unit is based on the most up-to-date survey/SUF information available, whereas the sampling stratum is based on the information that was on the SUF when the unit first appeared. Thus, the first-level imputation groups are:

$$\text{province}^* \parallel \text{KOB}^* \parallel \text{ZONE}^*,$$

where the values for province*, KOB* and ZONE* come from the most up-to-date information.

- 2) First-level imputation groups are combined into second-level and higher imputation groups (i.e., more aggregated) differently in the new imputation system, so that the units that are grouped together are more similar. For example, in the old imputation system, Québec units were grouped with Alberta units at the second level. This grouping does not occur in the new system until the fifth (and final) imputation level where units from across the country are combined. Different-sized units within a province are combined first. One practice that remains unchanged is that units from different KOBs are never combined in an imputation group.
- 3) Ratio imputation methods based on calendarized revenue from GST files and on GBI (gross business income recorded on the BR) are now used. For example, in ratio GST imputation, the mean sales of a group of acceptable donors divided by their mean calendarized GST revenue gives a ratio. This ratio is then multiplied by the recipient's calendarized GST revenue to obtain the GST-imputed value for sales. Ratio GBI imputation follows exactly the same procedure and is now used as the last-resort imputation option, instead of mean imputation that was in the old "last-resort" imputation method. Mean imputation used the mean current sales of all acceptable donors in the same imputation group as the imputed value for the missing record.

Increased Use of Administrative Data with the Implementation of the GST Model (Goods and Services Tax Data)

The MRCTS is one of the survey programs for which the use of existing administrative data (tax data) was tested. Although it was determined that direct replacement of survey data by administrative data was not feasible for this monthly survey, a somewhat more sophisticated model employing tax data as an auxiliary variable was found to be effective. 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 May 2004 reference month, the MRCTS will use both survey data and data modeled from GST returns to estimate monthly sales levels. The following is a brief description of the methodology tested and eventually used in this process.

About the 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 Canadian 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, while businesses with revenues between \$500,000 and \$6 million per year must remit quarterly, and 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.

Potential Replacement of Survey Data by GST Data

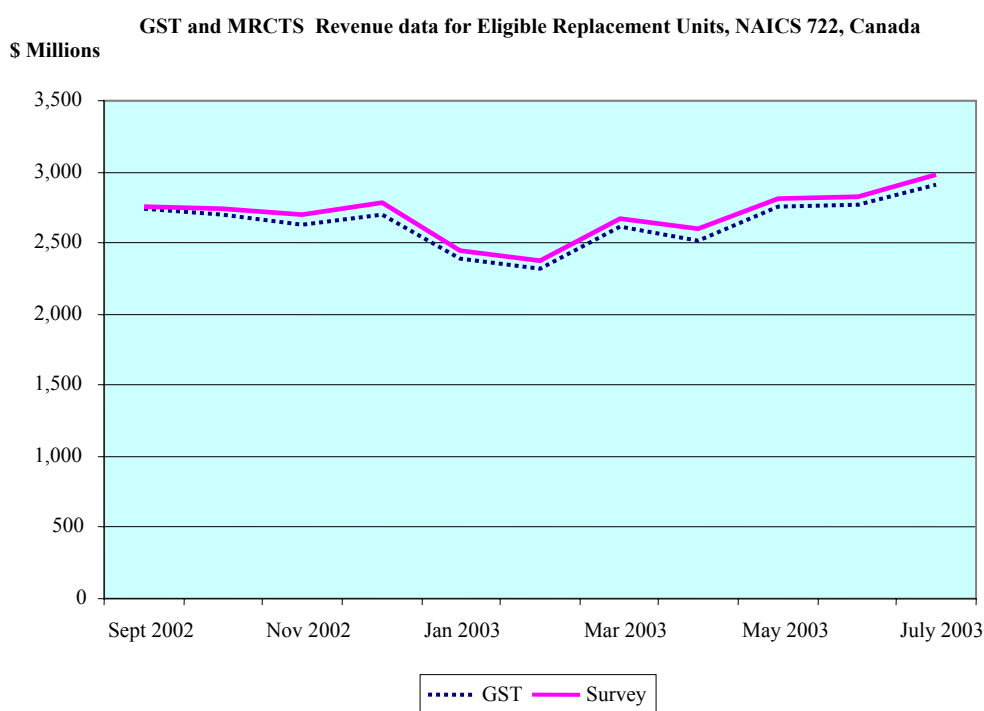
The MRCTS sample contains two types of units: simple and complex. For complex units (units consisting of more than one establishment), often only one GST value may cover all establishments. Consequently, use of the GST does not allow for breakdowns by province and / or NAICS. Thus, only simple units (units that represent only one establishment) were deemed potentially eligible for replacement by GST data.

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 significantly large (in terms of sales) in a particular NAICS industry in a particular province are available for replacement. This ensures that large or dominant establishments, even those classified as simple, will continue to be surveyed. Consequently, only “take-some” units were eligible for GST replacement.

Units selected for GST replacement would still be a part of the MRCTS sample and would contribute to MRCTS estimates in exactly the same way. The only difference would be that, for these units, values generated by a model that utilizes auxiliary GST data would be used in place of the collected data. In practice, this is very similar to regular survey imputation where auxiliary survey data is utilized to generate values to fill in for data that were, for one reason or another, not collected.

Studies were performed to determine the feasibility of using auxiliary GST data. The graph below shows the effect of replacing a certain portion of collected MRCTS sales with the values obtained from a GST model. These results were at the Canada, total KOB (NAICS 722) level.



The results of studies at provincial, KOB, and, in particular, province / KOB domains were considerably more variable. Indeed, it was found that the stronger correlations (at generally above 80%) are associated with the larger strata (NAICS, by province, by size). As a result, only large and medium sized “take-some” strata in NAICS 72211 and 72221 in the provinces of Quebec, Ontario, Alberta and British Columbia were selected for eligibility for GST replacement. Additional criteria were imposed to optimize the linkage between the MRCTS replacement units and the GST, to remove outliers, and to ensure there exists a reasonable monthly correlation between the MRCTS sales and the calendarized GST.

The GST model was initially implemented for the reference month of May 2004. However, the use of auxiliary GST data in MRCTS actually started one month earlier and has a wider scope than just the GST model itself. As described earlier, the GST is now used in regular MRCTS imputation. The use of the GST auxiliary data in the model and in the regular MRCTS imputation is virtually identical.

Starting with the May 2004 reference month, approximately 250 sample units were replaced with GST modeled data rather than surveyed. The number of units chosen for GST replacement was initially small due to the stringent criteria imposed upon the units used. However, the replacement of survey data with GST data has proven to conserve the high quality of the MRCTS estimates while reducing respondent burden for small businesses in particular. 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.

Impacts of New Methodology on Estimates and Time Series

To summarize, there are several factors that have an impact when comparing the old MRCTS series of estimates (ending in April 2004) and the new series of MRCTS estimates (beginning in April 2004):

- 1) Restratifying based on the updated April 2004 SUF and a new sampling procedure should improve the quality of the estimates and offset the fact that the total sample size has been reduced (2,900 to 2,250 sample units found to be live and in-scope by the survey). The impact from restratification alone, when comparing the “old” April 2004 estimates with the “new” April 2004 estimates, is estimated to be approximately 5% at the Canada level for all KOBs (i.e. overall) combined.
- 2) A backlog of new units (which should have been birthed in previous months) was added to the BR SUF in February 2004. They represented approximately 3% added to the total MRCTS frame revenue. These units should have been birthed onto the SUF prior in the months prior to 2004, instead of all in February 2004. Some units could actually have commenced operations prior to 2003. Therefore, the year-over-year increases from the 2003 overall estimates to the new overall 2004 estimates could be overstated by up to 3%.
- 3) The MRCTS imputation system was updated. Thus, even if the sample were to remain exactly the same, slightly different estimates would be obtained solely because of the change in imputation methods. The new imputation system should improve the quality of the estimates. The estimated impact from the new imputation system alone is approximately -2% overall. The combined effect of the restratification and the improved imputation at the Canada level, therefore, is 3%.
- 4) The May 2004 implementation of the GST model has had a negligible impact on the estimates (0.2% overall). The number of units for which the GST model is initially used was small due to stringent criteria on the units used. Still, the collected sample was reduced by approximately 250 units. More units will be added to the modelled portion in subsequent months, which will decrease both respondent burden and collection costs even more.

Users of MRCTS data should be aware of the changes related to the April restratification and the new methodology and should take these into consideration when comparing the new series (commencing with April 2004) with the old unrevised series (previous to April 2004).

Revisions to Historical Time Series

Background

The MRCTS sample on a NAICS basis was last drawn in January 1999. This sample has been maintained over the years following a standard process of removing deaths (units no longer in business) and adding births (new units added to the BR). However, after a period of time, this type of sample maintenance can lead to lower efficiency in sampling (in that we collect more records than would be the case if a full sampling were to be redone and, therefore, adding unnecessary response burden) and a higher CV (a CV, or coefficient of variation, is a measure of the sampling error). It is a general practice with monthly surveys to do a restratification and rebalancing of the sample within five years.

A restratified sample with maximum overlap with the old sample and re-balanced weights was drawn in April 2004. The new sample better reflects the new composition of the industry as of April 2004. The restratification, combined with the implementation of replacement data from the GST model for some sample units, reduced the size of the collected sample by approximately 30% (from roughly 2,900 units to 2,000 units). While reducing the overall sample size, the new sample has a higher quality (i.e., lower CV). An overlap period of one month occurred whereby both the old sample and new sample were collected and were used to produce estimates. This overlap afforded the opportunity to assess whether the new sample was indeed appropriate (and better). The CV at the total KOB-Canada level has improved, from 3.1% with the old sample to 2.3% with the new sample. The overlap period also enabled a link between the old series (April 2004 and prior) to the new series (April 2004 and after).

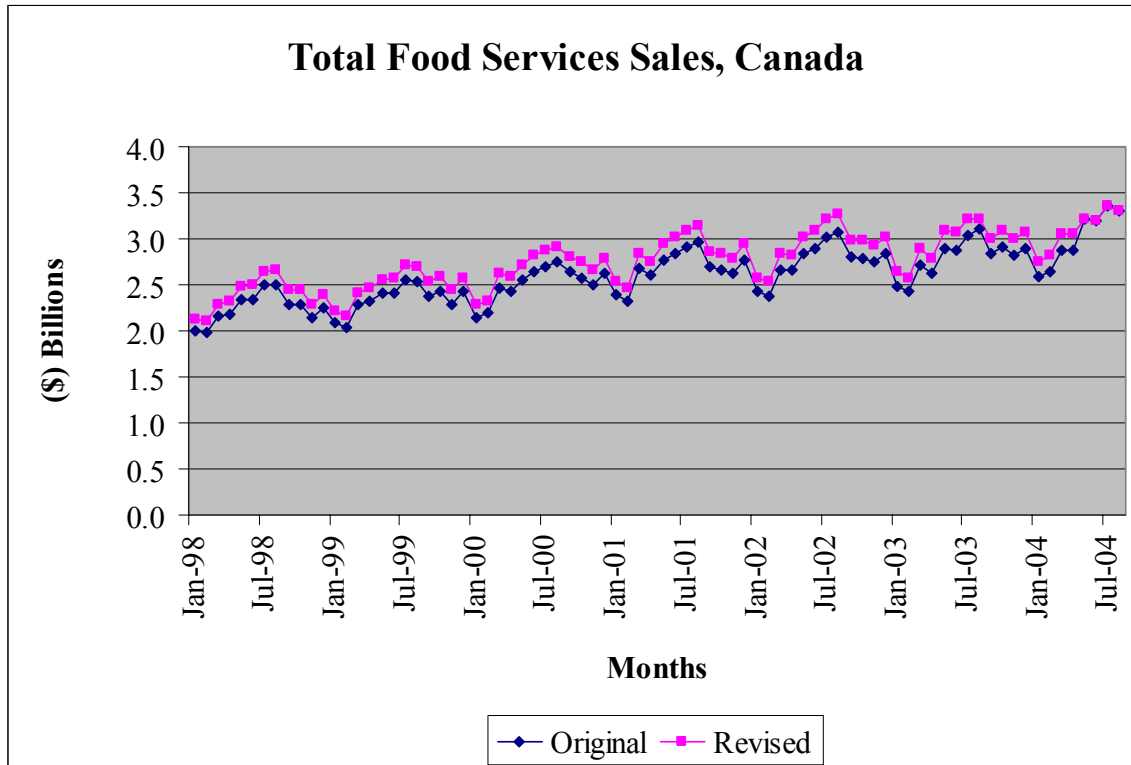
Revisions Process

Although a restratification with maximum overlap and weight rebalancing will, in general, minimize differences in estimates, it does not mean that the resulting estimates from the two samples will be identical. Indeed, one of the reasons for the sample update was to better reflect the current population on the BR frame and, accordingly, in the economy. The table below shows the relationship (link factors) between the two April samples.

Sales Estimates, Total Food Services, April 2004, \$'000

	Apr-04 (Old)	Apr-04 (New)	Link
Canada	2,881,510.2	3,048,347.0	0.9453
Newfoundland and Labrador	29,886.2	34,565.3	0.8646
Prince Edward Island	10,641.7	11,099.9	0.9587
Nova Scotia	84,183.3	76,211.3	1.1046
New Brunswick	58,774.4	56,414.7	1.0418
Quebec	562,527.1	641,921.2	0.8763
Ontario	1,090,255.7	1,229,499.5	0.8867
Manitoba	79,663.2	83,585.8	0.9531
Saskatchewan	77,237.1	81,979.7	0.9421
Alberta	363,989.6	348,911.3	1.0432
British Columbia	512,710.7	475,341.4	1.0786
Yukon	3778.2	2,391.1	1.5801
Northwest Territories	7,327.0	5,836.4	1.2554
Nunavut	536.0	589.4	0.9093

In order to provide users with a consistent time series back to January 1998, a simple linking approach was used. As a first step, the new sample level for April was applied to the old series' month-to-month movements back to January 1998. The starting point for the revisions process was the Total Food Services (total KOB) for each of the provinces/territories. This resulted in changed levels and monthly movements at the Canada level. As well, additional revisions were made to reflect frame or data errors not previously incorporated in the published data. Finally, annual movements in the calendarized GST data were also used to guide and verify the revisions. As seen in the Chart below, the method used to revise the historical series endeavours to keep the historical year-over-year and month-to-month movements, for the most part, unchanged; the most notable change, therefore, is that the historical levels are higher than previously published.



Future Plans for the MRCTS

Further to the methodological changes discussed in this report, and the resulting revisions made to the historical series at the total KOB level for Canada and the provinces / territories, anticipated modifications to the MRCTS include the following:

- Evaluate and implement revisions to the historical series at the detailed KOB levels.
- Harmonize the concepts between the Annual Survey of Food Services and Drinking Places and the MRCTS.
- Benchmark the MRCTS to the Annual survey.
- Implement seasonal adjustment for the MRCTS.