

ESTIMATING MONTHLY SALES BY COMBINING SURVEY DATA WITH ADMINISTRATIVE DATA IN THE MONTHLY SURVEY OF FOOD SERVICES AND DRINKING PLACES

François Brisebois and Stuart Pursey¹

ABSTRACT

The Monthly Survey of Food Services and Drinking Places of Statistics Canada provides estimates of monthly sales and the number of locations. Recently this survey has been redesigned and a major element of the redesign is the extensive use of administrative data. The goal of this paper is to describe the approaches used to combine collected sample survey data from this survey with data from the Goods and Services Tax (GST), thus providing a way of maintaining data quality while reducing survey costs and reducing the burden on respondents.

KEY WORDS: Administrative data, Goods and Services Tax, Ratio estimation.

RÉSUMÉ

L'Enquête mensuelle sur les services de restauration et débits de boissons de Statistique Canada fournit des estimations des ventes mensuelles et du nombre d'emplacements. Récemment, cette enquête a subi un remaniement dont un des éléments majeurs est l'utilisation étendue des données administratives. L'objectif de cet article est de décrire les approches utilisées pour combiner les échantillons de données d'enquête recueillis avec les données de la taxe sur les produits et services (TPS), et ainsi fournir un moyen de maintenir la qualité des données tout en réduisant les coûts liés à l'enquête et le fardeau du répondant.

MOTS CLÉS : données administratives; estimateur par quotient; taxe sur les produits et services.

1. INTRODUCTION

1.1 Overview

The Monthly Food Services Survey (MFS) publishes estimates of the monthly value of sales and the number of locations by province / territory for food services industries based on the North American Industry Classification System (NAICS):

- Full-service restaurants (NAICS 7221);
- Limited-service restaurants (NAICS 7222);
- Caterers, Food Service Contractors, and Mobile Food Services (NAICS 7223);
- Drinking Places (NAICS 7224)

The MFS is a replacement for and redesign of the Monthly Restaurants, Caterers and Taverns Survey (MRCTS). The redesign took place during 2006 and 2007 and on February 1, 2008 the first preliminary estimates for November 2007 were released in Statistics Canada's Daily. The new survey was run in parallel with the old survey from March to October 2007 to validate results and assess the changes brought by the adoption of new approaches.

The purpose of this report is to describe the strategy of the methodology of the redesign. There is emphasis on the use of administrative data and the way in which it is combined with survey data. This section provides background to the MFS and its redesign. Section 2 describes the Business Register and the use of the Goods and Services Tax (GST)

¹ François Brisebois and Stuart Pursey, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa ON K1A 0T6
François.Brisebois@statcan.gc.ca Stuart.Pursey@statcan.gc.ca

administrative data at Statistics Canada. Section 3 discusses the methodology of the redesign and section 4 provides outcomes to the redesign and continuing challenges.

1.2 Goals of the redesign

The main elements and goals of the redesign were set in 2006. They are listed below with some explanations.

- Improve the target population to cover businesses that are either employers or non employers (the old MRCTS covered only businesses that were employers).
- Restratify the MFS frame every two years and select a new sample.

This strategy helps to moderate the effects of the volatility in the MFS population. Thus every two years the MFS has a new sample selected from the most current BR. However the previous sample continues on for six months into the new 24 month period to provide a parallel (old and new sample) basis for monitoring the impact of the new sample on the estimates. As a historical note, the initial sample of the redesign was selected in March 2007. Then in December 2008 the first restratification will be performed, beginning the 24 month cycle.

- Reduce response burden by introducing take-none strata and by extending the use of administrative data.

Traditionally, administrative data has been used in Statistics Canada business surveys as an alternative, substitute, or replacement for collected survey data—the strategy has been to incorporate administrative data within a traditional business survey design. Instead, in the MFS design the goal is to use two approaches. For one part of the population derive a traditional sample design based on collected survey data. For the other part, develop a design that combines GST administrative data with survey data.

- Reduce response burden and make better use of collection resources by implementing more data collection agreements for the chains.

The MFS has negotiated reporting arrangements with several major restaurant chains to collect data for all of its outlets through the head office. Thus the Head Office becomes the sampling unit and collection entity, rather than the individual outlets. This approach decreases response burden since outlets belonging to the chain are not subject to sampling. Also it frees some of the MFS's fixed sample size for assignment elsewhere (since the maximum number of units sampled for the chain becomes one rather than something potentially much greater than one).

As well, deriving data from all of the outlets as a single collection unit with a single contact point (generally) improves data quality—in effect it is a census with no sampling error of the outlets belonging to the Head Office.

It is awkward in the survey process to introduce new reporting arrangements within a restratification cycle; for that reason strong efforts are made to negotiate new reporting relationships so that they are prepared and ready to be used for the next upcoming restratification.

- Produce seasonally adjusted data. Data were backcasted from August 2007 to January 1998 to take into account the methodological changes introduced with the new survey.

1.3 Challenges in estimation in the food services industry

There are difficult challenges to overcome in producing estimates for the food services industry:

- There is lots of volatility in the food services industry: many births and deaths from month to month and many changes in ownerships; thus the identification of changes from month to month is important.

- The chains have a large impact on the estimates and therefore finding, recognising, and identifying the elements (locations) of chains helps to improve the quality of data collection by developing reporting arrangements which in turn reduces response burden significantly.
- Accuracy and stability of estimates:

The change in level from month to month is an important and visible indicator and thus this measure must be as accurate as possible given methods, resources and the data confrontation of subject matter analysis.

In month *m* the preliminary estimate is published for *m-2*, the first revision is published for *m-3*; the second revision is published for *m-4*. There will be an annual revision but its timing is not yet set. Each revision should be an improvement in accuracy to the previous: there must be minimum volatility in the revisions caused by methods.

2. An overview on Statistics Canada’s Business Register and the use of administrative data

2.1 The Business Register

The Business Register (BR) is a list of businesses engaged in the production of goods and services in Canada. At Statistics Canada the BR is used to create frames from which sample designs can be derived. There are about 2.3 million alive and active businesses on the BR. Almost all (slightly more than 99.1%) are businesses so simple in their structure, characteristics, and data availability that they are best described within a single record in the BR. The remaining 20,000 businesses are more complex: the operating and legal structure of their production units is derived through in-depth interviews with the businesses. The units of a complex business are arranged in a multi-level hierarchical pattern that describes their reporting relationships and data availability.

Early in the 1980s Statistics Canada developed a standard statistical model that defines enterprise-based and establishment-based surveys. (Also, there are company-based surveys and location-based surveys, but since they are rarely used they are not described here.) *Enterprise*-based surveys measure the financial position and performance of businesses. *Establishment*-based surveys measure the revenue and expenses in the production of goods and services. Thus a key goal of a survey design is to identify and contact the part of a complex business that is best-suited for reporting data. In the multi-level hierarchical pattern of a complex business the enterprise is at the top: it is the “parent” of the units below it: there may be (and often are) several establishments. But in a simple business, because it is a single unit, that unit represents both the enterprise and the establishment at the same time. In the MFS, both the enterprise and its establishments are important in deriving the MFS sampling unit and constructing the MFS frame.

Close to 84,000 establishments are classified under NAICS 722. Of them, 7,500 establishments are linked to about 1,100 complex enterprises (and thus the remaining establishments are simple businesses). There is another type of business that is usually, but not always complex. These are the chains: a group of businesses related within a franchise or chain agreement. Where possible the MFS derives a data collection relationship with its Head Office.

2.2 Using administrative data at Statistics Canada

Statistics Canada has had access to administrative data since the 1970s. There are numerous agencies in Canada that provide administrative data to Statistics Canada. For the sub-annual business surveys, the most important source is the Goods and Services Tax (GST) which is provided by Canada Revenue Agency (CRA) under a memorandum of agreement that protects the security and privacy of tax data.

GST data is like all sets of data whether administrative or survey: there are difficult issues of timeliness, accuracy, and coherence to resolve. In Statistics Canada, Tax Data Division and Business Survey Methods Division are centres of expertise that manage these issues. The following is a list of the most important characteristics of GST data that are addressed (Pierre, 2005).

- The GST provides a monthly estimate of sales. It takes about 3 to 18 months to fully populate the data of a particular month and thus the accuracy of imputation is important since each imputed datum is in effect a temporary substitute for its forthcoming GST datum.
- Businesses that report GST do so monthly, quarterly, or annually. For those that do not report monthly an approach using calendarization methods is used to derive data on a calendar month basis.
- GST data are available for the legal entity. Since for simple businesses, the legal entity and the establishment are represent the same unit, GST data (after dealing with the issues mentioned above) can be used as a replacement for survey data. There are efforts to extend the use of GST data to the complex businesses and they are in development.

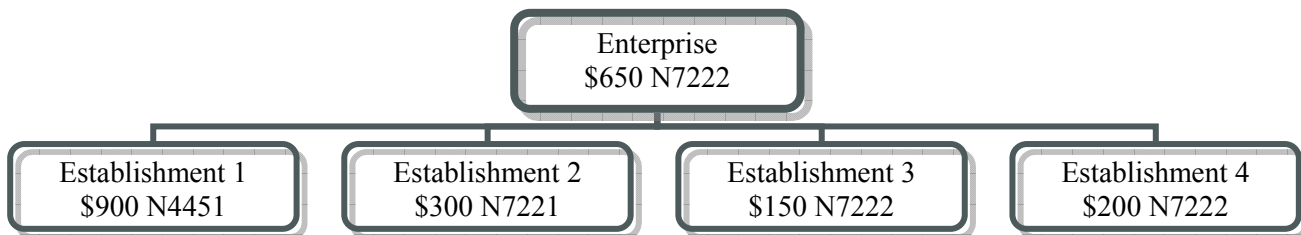
3. The methodology of the MFS

3.1 The MFS frame and sampling unit

Each month the Business Register Division produces a Generic Survey Universe File (G-SUF) from the current BR. The MFS frame is a subset of the G-SUF: it contains the statistical units of the G-SUF that are flagged as establishments, classified as alive and active, have a NAICS code that begins with 722, and have a non-zero revenue measure of size.

The sampling unit of the MFS is the cluster of in-scope establishments that have a NAICS beginning with 722 and belonging to the same enterprise. In practice the cluster is almost always the Enterprise itself.

There are some enterprises that contain establishments that are not NAICS 722. In these cases the revenue of the cluster is the sum of the revenues of the establishments that are NAICS 722. The 4-digit NAICS of the enterprise is the dominant 4-digit NAICS of the NAICS 722 establishments, based on summing revenues by 4-digit NAICS. The example below shows the data of an enterprise—the MFS sampling unit—after it has been modified to represent the cluster of establishments belonging to NAICS 722.



3.2 The MFS sample design

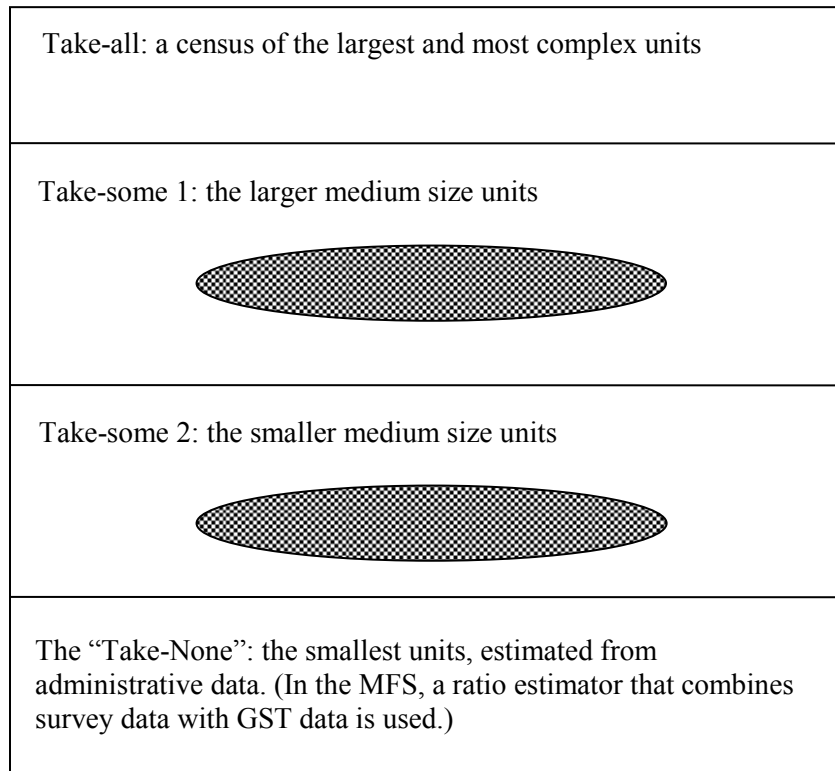
The MFS releases estimates by province / territory and 4 digit NAICS, and thus the first task is to derive these 52 strata. Given this beginning step, the strategy of the MFS redesign is to separate the population of the food services industry into two sub-populations.

For one part of the population a traditional sample design approach is used. This includes the enterprises in the territories and the Maritime Provinces for NAICS 7221, 7222, 7223, and 7224 and in British Columbia, the Prairie Provinces, Ontario and Quebec for NAICS 7223, and 7224.

In this traditional approach (see Diagram 1) there is a take-all stratum of the largest and most complex businesses (this includes the chains, whether large or small). There is a take-none stratum of the smallest business where there is no sampling and instead estimation is accomplished using administrative data. Usually many, many sampling units are in the take-none and thus collection costs and response burden are much reduced. The take-some strata include the medium size businesses and often there are two of them. The boundaries between the take-all and take-some are determined to be optimum so as to derive a minimum sample size for the target CVs.

For the take-none strata the ratio estimator that integrates GST data with survey data is used: see section 3.3. For the territories there are no take-some strata because the population is so small. Instead there is just a take-all stratum and a take-none stratum.

Diagram 1: the traditional business survey sample design used at Statistics Canada



Section 3.3 An approach that combines administrative data with survey data

One of the goals for the MFS redesign is to increase the use of administrative data. It has been essential to understand the relationship between survey data and administrative data. Efforts have been made to derive sample designs and estimators that integrate administrative data with survey data. Dubreuil et al (2003) and Hidiroglou et al (2003) are two papers that discuss and evaluate estimators for sub-annual surveys that use GST data. The ratio estimator emerged as the best approach, with cautionary advice to use robust estimation by removing outliers, avoiding strata with low counts, and using the estimator in strata where the correlation between GST data and survey data is reasonable.

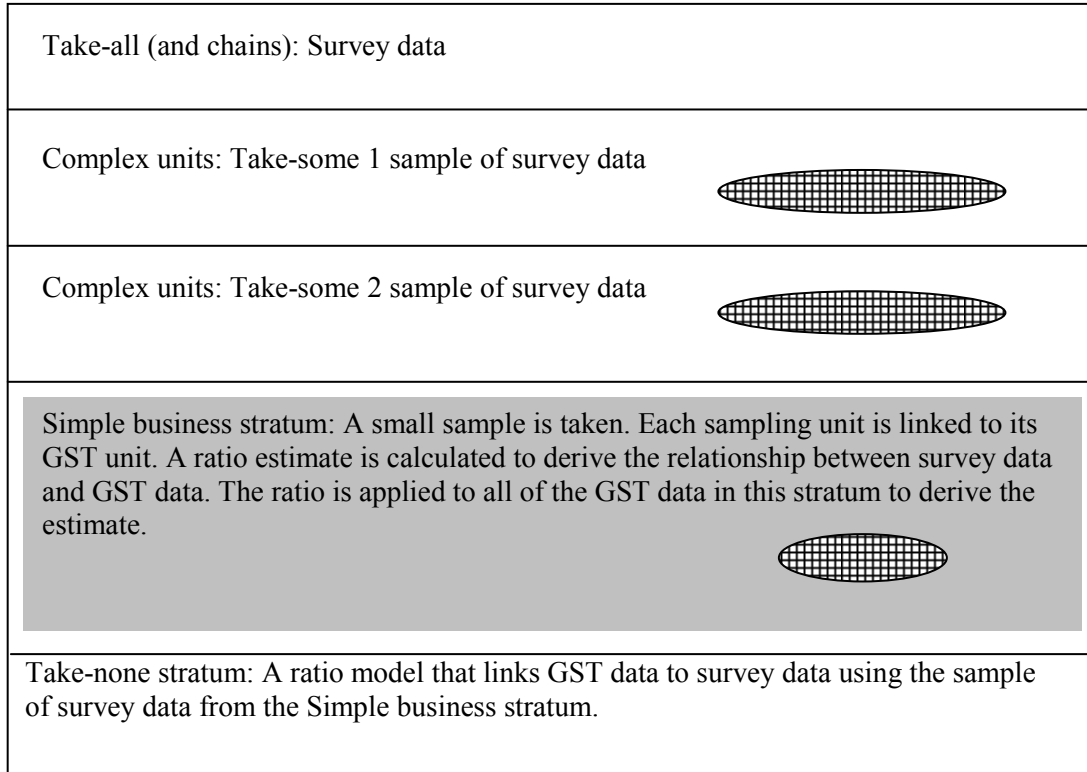
There are two aspects to the use of the ratio estimator in the MFS. For the preliminary estimate, for reference month m , the ratio estimator bridges GST data from month $m-1$ to the collected data of month m ; that is, it adjusts for the economic trend from the previous to the current month. (As well it adjusts for conceptual differences between the two sources of data.) For the revisions that follow, GST data and collected survey data are available for the same month and thus the ratio estimator adjusts for the conceptual differences.

In the MFS, the approach of combining administrative data with survey data is used for the simple businesses within British Columbia, the Prairie Provinces, Ontario and Quebec for NAICS 7221, and 7222. Also, as mentioned above, it is used for the take-none strata throughout the MFS. Diagram 2 below illustrates the strategy. There is a take-all stratum (including the chains) of the largest and most complex units and there are two take-some strata of complex units. For these, collected survey data is used.

The remaining units are simple businesses: they are assigned to the simple business strata (about 24,000 units) and the take-none strata. For these strata a relatively small sample of collected survey data is taken (about 700 units) and matched

to their corresponding units with GST data. The ratio estimator is calculated from the matched sample and then applied to all GST data within the simple business strata and the take-none strata to derive the MFS estimates.

Diagram 2: The ratio model approach



4. Outcomes of the redesign and challenges

The redesign has been successful in reducing response burden and collection costs and still maintaining quality. For perspective, the old MRCTS used data from GST for 13% of the frame (11,000 establishments), which represented about 20% of total sales. In the redesign, the MFS derives data from GST for 78% of the frame (65,500 establishments from the take-none and simple business strata), which represents more than 50% of total sales. Many more data collection agreements have been negotiated and efforts continue to make more.

There are several challenges for the near future. These include revising the method of calculating response rates to include the impact of GST administrative data. Also, deriving a stable and accurate ratio estimator remains a challenge due to the volatility of the food services industry. Finally, there is development work with the objective of deriving an estimate of the variance of the estimate of month-to-month change.

REFERENCES

Dubreuil G., Hidioglou, M.A, Pierre L. (2003) “Use of administrative data in modeling of monthly survey data”. Proceedings of the Survey Methods Section, SSC Annual Meeting, June 2003

Hidioglou, M. A., G. Dubreuil, G. and Crowe S. (2003) “Using the Goods and Services Tax (GST) for Monthly Surveys: Application to the Monthly Restaurants, Caterers, and Taverns Survey”. Internal Statistics Canada document

Pierre, Louis (2005) “Use of GST data in Statistics Canada sub-annual surveys”. Internal Statistics Canada document