

QUARTERLY SURVEY OF FINANCIAL STATISTICS FOR ENTERPRISES

Pierre Daoust, Nora Bohossian, Sylvie Gauthier, Geneviève Demers, Jill Nicholas¹

ABSTRACT

The Quarterly Survey of Financial Statistics (QFS) is Statistics Canada's principal source of current information about the incorporated private business sector. The survey provides estimates of standard accounting variables including profits, balance sheet accounts and savings and investments in Canada. This information is used by governments, banks, and other institutions, and provides an essential input into the Canadian System of National Accounts. The survey design includes a take-all stratum of the largest corporations, and a sample of medium sized corporations. The results are reconciled with the Annual Financial and Taxation Statistics figures, a census of financial data available from corporate income tax returns. In this paper we discuss the methodology of the QFS and future survey improvements.

KEY WORDS: Financial Survey Methodology; Stratification on two variables; Take-all Component Definition; Take-none Component Estimation

RÉSUMÉ

L'enquête trimestrielle sur les statistiques financières des entreprises (SFT) est la source principale de l'information la plus récente à Statistique Canada en ce qui concerne le secteur des entreprises privées constituées en société. L'enquête publie des estimations sur les variables normalisées de la comptabilité incluant les profits, les comptes et épargnes des bilans et les investissements au Canada. Cette information est utilisée par le gouvernement, les banques et les autres institutions, et fournit un intrant essentiel pour le Système de comptabilité nationale du Canada. Le plan de l'enquête comprend des strates à tirage complet des plus grandes entreprises et un échantillon des entreprises de taille moyenne. Les résultats sont réconciliés avec les chiffres des Statistiques financières et fiscales annuelles (SFFA), un recensement des données financières disponibles à partir des déclarations de revenus d'une entreprise. Dans le cadre de cette présentation, nous discuterons de la méthodologie de la SFT et des améliorations futures de l'enquête.

MOTS CLÉS : Définition composante à tirage complet; estimation composante à tirage nul; méthodologie d'enquête financière; stratification à deux variables.

1. INTRODUCTION

The Quarterly Survey of Financial Statistics (QFS) is part of the Statistics Canada Financial and Taxation Statistics program for Enterprises. Data from the program are published quarterly through the QFS, and on an annual basis through the Annual Financial and Taxation Statistics (AFTS).

¹ Pierre Daoust, 100 Tunney's Pasture Driveway, Ottawa, Canada, K1A 0T6, pierre.daoust@statcan.gc.ca; Nora Bohossian, 100 Tunney's Pasture Driveway, Ottawa, Canada, K1A 0T6, nora.bohossian@statcan.gc.ca; Sylvie Gauthier, 100 Tunney's Pasture Driveway, Ottawa, Canada, K1A 0T6, sylvie.gauthier@statcan.gc.ca; Jill Nicholas, 100 Tunney's Pasture Driveway, Ottawa, Canada, K1A 0T6, jill.nicholas@statcan.gc.ca; Geneviève Demers, 100 Tunney's Pasture Driveway, Ottawa, Canada, K1A 0T6, genevieve.demers@statcan.gc.ca

The QFS is the main source of current information concerning corporate profits in Canada. This enterprise²-based survey of the corporate sector is designed to obtain information on corporate income statements (elements of revenues, expenses and profits) and balance sheets (assets, liabilities and equity). The QFS data include several hundred of the standard accounting variables. While they are published on a quarterly basis to provide a current view of the health of the corporate economy, they are important as a major input into the quarterly Gross Domestic Product (GDP) estimates, measures of wealth and sources and uses of funds for investing. In addition to private sector economists and the Canadian System of National Accounts, the QFS data are used by other government departments, particularly the Department of Finance, Industry Canada and the Bank of Canada.

The data used to produce the AFTS is compiled from a census of corporate information. Such a census is made possible by supplementing data collected through the QFS statements and the survey of provincial and federal level government business enterprises, with administrative data available from the Canada Revenue Agency (CRA), Canada's central taxation authority. The annual financial statistics comprise information similar to that in the quarterly program. They are available with greater industrial detail and also include corporate tax information not available from the quarterly survey. The annual corporation financial statistics aggregates are reconciled with estimates from the quarterly survey statistics. This function is important since the quarterly survey has no sample coverage for the small corporations and depends entirely on the annual estimates in this area.

This paper discusses the methodology of the QFS and future survey improvements. Section 2 provides the reader with a general description of all the survey steps. Section 3 discusses challenges associated with the survey methodology and proposes improvements. Future considerations and conclusions follow in Section 4.

2. SURVEY DESCRIPTION

2.1 Survey Frame, Population and Sampling Scheme

The QFS collects data on the for-profit incorporated businesses sector in Canada for eighty industry groups in the economy. The QFS population is the largest component of the business sector, one of the three main sectors of the domestic economy in the Canadian System of National Accounts. The other two sectors, the individuals sector and the government sector, are excluded from the QFS.

The QFS frame is extracted from Statistics Canada's Business Register (BR) (Statistics Canada, 2007), a complete listing of units engaged in economic production within Canada. The BR represents businesses through their legal and operating structures. The QFS is an enterprise-based survey. Complex enterprises (consolidations) generally produce consolidated financial statements and these structures are used as the QFS sampling and collection entity.

The total QFS population size exceeds one million enterprises. It is stratified by size using two size stratification variables, revenue³ and assets, and by industry based on the aforementioned 80 industry groups.

In general, within each industry group, large businesses form a take-all (TA) stratum, medium businesses form a take-some (TS) stratum and small businesses form the take-none (TN) stratum. All businesses above a designated size threshold are selected for the TA stratum, a sample is selected for businesses falling within the TS thresholds, and none of the businesses falling below the exclusion threshold of the TN stratum are surveyed. Five industry groups consist only of a TA stratum (that is, they are census industries) and nine industry groups do not have a TS stratum (that is, each business classified under this industry group either belong to the TA stratum or the TN stratum).

The boundaries of the TA stratum for each industry group are derived in the first quarter of each year and are kept fixed for the remaining quarters of that year. An iterative algorithm is used to derive these boundaries; an adaptation of the Lavallée-Hidiroglou algorithm (1988) in the context of two auxiliary variables. The algorithm determines the TA boundaries that would yield the minimum required sample size while satisfying pre-specified constraints on the coefficient of variation of total revenue and total assets estimates. The sample size in each TS stratum must also satisfy a

2 QFS defines an enterprise as the highest legal unit in a corporate structure that can provide complete (consolidated) financial statements. It should be noted that this might differ from the definition of an enterprise used outside of Statistics Canada.

3 References to revenue for QFS generally correspond to the operating revenue of the enterprise or consolidation. However, for smaller simple businesses, the revenue used for the stratification is extracted from tax data and corresponds to gross revenue.

pre-specified minimum sampling fraction and a minimum sample size of four enterprises. Finally, a special pre-determined TA boundary for revenue is used with respect to complex enterprises (consolidations).

A stratified simple random sampling strategy is used to select the survey units. The first quarter sample moreover is selected to maximize the overlap with the sample of the previous year. The Kish and Scott (1971) method is applied to achieve this objective. For the subsequent three quarters, the sample is updated by removing enterprises that are no longer active (deaths), adding a sample of enterprises that are new to the quarterly population (births), and adding new enterprises that fit the TA criteria. The total sample consists of roughly 5,500 enterprises.

2.2 Data Collection

The survey questionnaires comprise financial statements typically prepared by incorporated businesses. Corporate activities across the economy are extremely diverse, resulting in the utilization of a variety of unique financial reporting variables. To accommodate the diversity in financial reporting across industries, fifteen different questionnaires are used. The majority of the sampled enterprises participate by returning their data by mail, fax, or electronically (Electronic Data Reporters, EDR). The number of EDR participants has continued to grow since the program was introduced in 2003. For the first quarter of 2008, roughly 20% of the sampled enterprises agreed to participate in the EDR program. Special collection arrangements are also made for a small set of respondents (for example, local credit unions and banks). Near the end of the collection cycle, non-respondents are prompted by fax to at least provide 3 Key Variables (3KV) by the survey processing deadline. However, this does not relieve the respondent of furnishing a full report later. The 3KV are assets, revenue and operating profit.

The response status of sampled enterprises is monitored through the use of the generalized software BLAISE (Statistics Netherlands), which is also used to schedule fax and telephone follow-ups. As is the case for many other Statistics Canada business surveys, a score function is used for the QFS (Daoust, 2007) to prioritize follow-ups. Score functions aim to allocate collection resources optimally in the presence of highly skewed populations. The QFS's score function prioritizes the follow-up of non-respondents by taking into account the businesses' contribution to the total revenues and total assets for the industry group. This score prioritizes the units in each industry group in order to achieve a targeted coverage for both assets and revenue. It also prioritizes special treatments for extremely important enterprises. The vast majority of the follow-ups done on non-respondents happen within a two-week window. Non-responding units with the greatest impact on estimates, as well as very large enterprises, are treated either by more senior staff or by a special team. Scores are produced multiple times for each quarter during the collection period. These represent dynamic processes that integrate the up-to-date status of sampled enterprises in collection activities.

2.3 Data Processing

The QFS data editing is shared by Statistics Canada's Operation and Integration Division (OID) and the Industrial Organization and Finance Division (IOFD). Most reporting and data entry errors are corrected as a result of the intricate computer capture and edit procedures applied to the data. The QFS utilizes trained accounting staff to capture and analyze reported data to minimize the frequency of non-sampling errors. This is particularly effective for financial data where accounting relationships are established and balancing is required. Analysts also review outlier values.

At the data processing stage, IOFD analysts can make modifications to the input file that will be used in the estimation process. These modifications consist of adding important units that were either not in the frame at the time of sampling or were not sampled due to incorrect size measures information (referred to as mega-adds), correcting NAICS classifications (referred to as mega-changes), and identifying any outlier unit which should have a weight of 1 (normally large TS unit).

Although significant effort and resources are used to maximize response rates, there is always an element of non-response. Historical imputation is used for total or partial (3KV response only) non-response where there is previous quarter data. Imputation for total non-response is based on the movement observed for reporting units in the non-respondents' industry. For partial response, the trends are computed using the 3KV's reported for that unit. For birth records, an attempt is first made to seed the 3KV where current auxiliary information is available and then proceed with partial historical imputation. For the remaining births, which tend to represent only a small fraction of the sampled enterprises, donor imputation is used based on a nearest neighbour strategy. In the case of complete non-response, the matching variables used are the frame values for revenue and assets, whereas the respondents' reported values are used in cases of partial non-response.

Note that “proc outlier” from Statistics Canada Banff system (Banff Support Team, 2008), a module used to detect outliers, is in place to omit records with large reported values for assets or revenues from the pool of donors.

The overall estimates are derived from two different components: the sampled portion and the non-sampled (TN) portion. For the sampled portion, the Statistics Canada Generalized Estimation System (GES Support Team, 2005) is used to compute the estimates and associated CVs. A post-stratification methodology is used to derive the estimates of totals per industry group, which are made available 58 days after the end of the reference quarter. The final weight of a unit is the product of its design weight times the calibration weight, commonly referred to as g-weight in GES, which is an adjustment based on updated population counts reflecting the current quarter.

As mentioned before, some units will see their final weight set to 1. These special units are: 1) units which have reported assets or revenue twice as large as the TA boundary, and are considered to be outliers, 2) mega-add units, and 3) very large TS units. Mega-changes may also affect final weights as units can be assigned a weight of 1 if they are above the TA boundaries of the new industry group.

For the TN portion, the estimate for each industry group is derived by applying the quarter-to-quarter movement of QFS sampled portion to the previous quarter TN estimate. As a result of the reconciliation process mentioned previously between the QFS and the AFTS, the previous quarter estimates represent essentially the AFTS aggregates for the corresponding TN component. More details on the TN estimation are provided in section 3.2.

The AFTS key aggregates are reconciled with the corresponding QFS estimates for most of the industry groups once the AFTS aggregates are published (roughly a year after the end of the reference year), and a year later once both of these figures have been revised. This yearly process is applied independently for the survey component (TA and TS population) and the TN population. The reconciliation process occurs initially at the annual level for fifty important variables common to both programs (QFS estimates are annualized for this first step). It is used to reconcile other QFS variables using the QFS fourth quarter weighted data, and finally quarterized in order to revise QFS quarterly estimates.

3. CHALLENGES FOR 2008/2009 AND BEYOND

3.1 Improvement to the sample design

The QFS sample design has evolved over the years under the guiding principle of achieving a reasonable compromise between the 1st quarter sample optimization and sample stability over time. In this section, three potential improvements that could impact these factors are discussed: 1) adjustment of the minimum sampling fraction and TA upper bound values 2) creation of two TS strata for some industry groups and 3) potential changes to stabilize the TA populations over time.

3.11 Minimum sampling fraction (MSF) and TA upper bound values (CAPS)

As discussed earlier, minimum sampling fractions (MSF) are incorporated into the sampling strategy to avoid large weights. These were established a few years ago, and were set to 10%, 12.5%, 20% or 25% depending on the industry. The TA CAP values are used in the QFS to ensure that key large enterprises are surveyed. The QFS data collected for these enterprises are also a component for other major statistical programs at Statistics Canada, and most of the CAP values currently used were determined in 2004 by considering their needs.

It has been observed for the last few years that the MSF and CAP values were frequently used, potentially creating unnecessary over-sampling. A study was conducted comparing the values currently used to optimal⁴ values under the sampling strategy. The objective of the study was not so much to reduce the sample size for the QFS, as it was to consider reallocating the gains achieved to either increase the sample size to target better precision for specific industries and/or to consider approaches that would improve the estimation of the TN population.

The first phase of the study analysed different combinations for MSF and CAP for previous quarters, considering several options lying between the current and optimal values. These options involved a few survey characteristics: gains (or losses) in sample size, impact on sample overlap between quarters (and in particular the number of new units to be pre-

4 Optimality is defined as the values determined by the QFS sampling methods prior to applying the MSF and TA CAPs.

contacted), and reduction (increase) in coefficient of variation (CV) values (frame data for revenues and total assets). CVs were also estimated from reported data in order to better evaluate their impact on the precision of the QFS estimates. This initial analysis identified several industries for which the MSF could be lowered and/or the TA CAP could be increased while still attaining the target CVs. The modification of the MSF was seen as potentially decreasing the sample size by roughly 10% in the targeted industries, mostly affecting TS units. On the other hand, changing the TA CAPS was seen as decreasing the sample size by 6%, with most of these impacting TA units dropping from the sample. Given the impact this had on the largest enterprises and the desire to coordinate any decision on TA CAPS with other financial surveys, it was decided to postpone the investigation on the TA CAPS until QFS has completed its integration into Statistics Canada's new Business Register environment

A subsequent study looked at the impact of changing the MSF on the quarter-to-quarter trends (revenue and assets). The standard errors of these trends were derived by adapting an approximation to estimating standard errors for trends developed in the context of rotating samples (Laniel, 1988). The results showed that for most industries targeted for a MSF revision, less than 20% increase in standard deviation was observed. The final decision on which MSF to use for Q1 2008 was based on the results from the two phases of the study, as well as response rates achieved over the last few years. A very conservative approach was adopted for industries with lower response rates. In order to reduce the risk of instability in quarter-to-quarter estimates associated with large weights, a decision was made not to drop the MSF below 0.10. Consequently, the MSF was lowered for some industries for Q1 2008 and sample size gain was reallocated to industries with increases in population size. The TA CAPS were not altered.

3.12 Creation of two take-some (TS) strata

The distribution of the TS stratum's population for some of the industry groups is highly skewed. Thus, we may find that for these industries, there are enterprises with very different revenue and asset values in the stratum resulting either in a larger sample size needed to attain the target precision or cross-sectional estimates that are variable when the sample size is set based on the MSF. It should be noted that changes to the sample over time with respect to deaths and births in the population and revisions to key design variables (in particular size measures) can cause more instability in the trend estimates. In addition, although many efforts are made to avoid this situation by considering several sources to update the BR, a portion of these population changes might be associated with the availability of information collected on sampled units (in particular out-of-scope units that are usually predominant in small enterprises). It is therefore desirable that the sampling strategy aims to detect and minimize this potential bias. Accordingly, for a potential efficiency gain, creating two TS strata instead of one seemed preferable if we wanted to have more homogeneous strata in order to be more efficient in allocating the sample. The work being done on this project is described below. More sophisticated analysis to determine the impact on final estimates (i.e. based on responded data) will be considered in the future.

Stratification to obtain the two TS strata is based on two variables: revenue and assets, the same variables used for the creation of the current single TS stratum. It was decided to preserve the strategy used to define the TA population. Although this might not be optimal for the two-stratum solution, it does ensure more stability of the key estimates while moving from one TS methodology to the two TS methodology. With the predetermined TAs, the stratification results are first obtained independently for each variable using the Lavallée-Hidiroglou method by specifying two TS strata. Any revisions to the TA boundary made by the independent processes are ignored, and the TS boundaries (revenue, assets) are combined to generate two TS strata: the first representing all TS enterprises that have a size measure above or equal to one of the TS boundaries, and the second representing all TS enterprises below both boundaries. The final sample size for each stratum is derived using Neyman allocation on both size measures, and by taking the maximum (still subject to MSF and a minimum sample size of 4).

A pilot test is being conducted in parallel with current production. The results obtained in the first few quarters of 2008 indicated that, for most of the industries for which the two TS approach is being considered, the CVs would decrease by 1 to 2 percent for roughly the same sample size (increased by roughly 1%). The sample size was actually slightly smaller for the two TS approach before applying the Kish and Scott overlap methodology (i.e. before overlapping with the Q4 2007 strata which were based on the one TS strategy). However, as could be expected, the sample size increased once the Kish and Scott method was applied as this was the first occasion on which the new stratification methodology was introduced. Under the new design, the number of new units in the sample increased by 15%, while the number of enterprises leaving the sample remained at the same level. This unusually large increase in the number of new units, impacting resources needed to carry the pre-contacts at the collection phase, would only occur the first time the new methodology was

introduced. The two TS Q1 2008 sample also seems less likely to be affected by sampling bias, based on results from a simple test that was conducted to determine whether a 95% confidence interval derived from the sample selected would cover true population totals. This is very positive and the approach seems promising. However, we need to continue monitoring the results closely to ensure that the estimates and trends remain stable in subsequent quarters.

3.13 Stability of the take-all (TA) stratum's population

A third improvement discussed in this section concerns the population stability of the TA stratum. Under the current methodology, enterprises can move from the TS stratum to the TA stratum on a quarterly basis, and for the first quarter can also move from the TA stratum to the TS stratum, which can result in instability in the estimates. This problem is even more prevalent in the first quarter, when the TA boundaries are recalculated for the new year.

We are considering changing the definition of the TA stratum to make it more difficult for enterprises to move in or out of the stratum. Alternatives that can be researched include the use of a buffer zone, generating the TA strata from data for more than one year, or extending the period for revising the TA boundaries. This challenge will become even more serious if the TA CAPS are increased to bring them closer to the optimal TA boundaries usually recalculated in the first quarter of each year to achieve the desired precision.

3.2 Estimation of quarterly trends for the take-none (TN) portion

Prior to 1999, the TN component was estimated by projecting forward the last reconciled QFS macro estimate (after reconciliation with AFTS). The projection strategy used trend ratios derived from QFS respondents common to the previous and current quarters. This approach is referred to as the ratio model.

For the 1999 QFS Redesign, a regression model was implemented for the TN portion. This regression model used historical data to model the relationship between the survey portion and the TN portion within each industry grouping for the 3KV. The parameters obtained formed a regression equation which was then used in production to estimate the TN portion of the quarter, using the survey estimates for the current quarter. Auxiliary information was also used as a quality tool to validate the level of the estimates. The remaining variables (other than the 3KV) were imputed using historical imputation. This approach was used from 1999 to 2002 and had several limitations, mostly related to the data conversion needed on key variables of the historical data and to adjustments needed on the projections. QFS returned to the simpler ratio model starting in 2003, lacking a better alternative to the unsatisfactory regression model approach, after a study demonstrated that these estimates were usually more precise than those derived by the regression model.

The ratio model, however, was initially replaced with the regression model as it was demonstrated that it is not in every industry that the small enterprises (TN portion) move in the same manner as the larger (survey portion) enterprises. Accordingly, the objective of improving the TN estimates remains important for QFS. Recently, research has started exploring other alternatives.

One alternative being investigated is the use of Goods and Services Tax (GST) data. The GST revenues refer to the sale of goods and services, investment income and other special elements of revenue. Although not all of the QFS population is covered by GST data, and the QFS revenue concept does not always correspond to the GST revenue concept, it is expected that it could be a reasonable proxy at least for some industries (e.g. construction and agriculture). Another alternative proposed would be to sample the TN population for some of the most problematic industries.

3.3 Improvement to Profits Estimates

Getting reliable and detailed information on profits⁵ is generally only possible when data are collected at the enterprise level (that is when complete financial statements are available). The QFS and the AFTS are major surveys at Statistics Canada designed to collect data at that level and for that purpose.

The QFS survey is designed to provide accurate estimates on total revenue and total assets. There is a generally poor correlation between these variables and profits for most industries. Despite this, the QFS provides relatively precise

⁵ References to profits for QFS and AFTS correspond to operating profits.

estimates of profits. Many features of the QFS design, among others the sampling of most of the large consolidations resulting in a dominant TA component for the survey, are key factors with respect to this. However, there are some obstacles to this approach. The first is the availability of reliable auxiliary information on the survey frame, with challenges usually relating to the timeliness of the data and to concepts of incompatibility. The second is the less-than-predictable nature of the profits compared to income and assets, as well as the rather dubious value of this information in terms of optimizing the survey design.

With the restructuring of the Business Register over the last few years, and with the other improvements made to the related data sources, this aspect will likely be reviewed. The improvements for the survey could be first considered at the sampling design phase, by using profits in the stratification strategy or over-sampling for specific industries. They can also be considered at the estimation phase, by considering using profits in the weighting strategy as an option to resolve timeliness issues involving the availability of reliable information on profits at the time of sample selection.

4. CONCLUSION

The QFS is a key statistical survey at Statistics Canada. The financial data released are key inputs to major statistics programs as well as to the Canadian System of National Accounts. Although the survey has been producing relatively reliable estimates since its 2004 redesign, a few possible improvements have been discussed in this paper.

Investigations on the MSF used for the TS component of the survey population and on the CAPS values for the TA boundaries have been conducted. They indicated that these parameters could be adjusted to allow for a more efficient sample allocation amongst industries, and to help reducing response burden for certain industries, while maintaining reliable estimates. Modifications have started to adjust the MSF, while modifications to the TA CAPS will be considered in coordination with other statistical programs. A study is being conducted to modify the stratification strategy of the TS component. One of the alternatives considered is creating two strata instead of only one stratum (as currently done) for a subset of industries. Indications are that this approach would create more homogeneous strata and improve the efficiency and stability of survey estimates. Research projects that consider alternatives to define the TA population are also envisioned to improve the stability of survey estimates across time.

Improvements are expected in the estimation of the QFS TN component, in particular by further considering using revenue derived from the Goods and Services Tax data for the estimation of the TN quarterly trends. Finally, the potential of incorporating auxiliary data on operating profit in the survey design, either at the sampling or at the estimation phase, to directly control or improve the precision for this key variable could be considered now that the QFS has been integrated into the improved Business Register.

The authors would like to thank Jack Lothian, H el ene B erard, Haig McCarrell and Jean Leduc for their key insights, which contributed to improving this paper.

REFERENCES

- Banff Support Team. (2008). "Functional Description of the Banff System for Edit and Imputation". *Statistics Canada Technical Report*.
- Daoust, P. (2007). "Prioritizing follow-up for the Canadian Quarterly Survey of Financial statistics for enterprises". *Papers presented at the International Conferences on Establishment Surveys (ICES_III)*, 1400-1407.
- GES Support Team. (2005). "GES v4.3 Overview". *Statistics Canada Technical Report*.
- Kish, L. and Scott, A. (1971) "Retaining Units after Changing Strata and Probabilities", *Journal of the American Statistical Association*, Volume 66, Number 335, 461-470.
- Laniel, N. (1988). "Variances for a rotating sample from a changing population". *Proceedings of the Business and Economical Statistical Section, American Statistical Association*, 246-250.

Lavallée, P., and Hidioglou, M. (1988). "On the Stratification of Skewed Populations". *Survey Methodology*, Volume 14, Number 1, 33-43

Statistics Canada. (2007). "Statistics Canada Business Register - A Brief Guide". *Statistics Canada Technical Report*.

Statistics Netherlands, BLAISE System, URL [http:// www.blaise.com](http://www.blaise.com)