27th Voorburg Group Meeting on Service Statistics

Warsaw, Poland
October 1-5, 2012-09-11

Use of Administrative Data in Statistics Canada's Business Surveys – The Way Forward

Wesley Yung and Peter Lys
Statistics Canada

Abstract

Statistics Canada has been successfully using administrative data in its business survey programs for several years now and more recently has begun to investigate ways to increase their use. For the most part, the current use of tax data have been integrated into existing surveys and systems which somewhat limited their use. However, two sub-annual surveys have recently been redesigned or developed which use significantly more monthly tax data. The Monthly Food Services Survey (MFSS) and the Quarterly Services Indicators (QSI) are using monthly tax data for a significant portion of units in the population. In the situation of the MFSS, a small sample of units for which tax data are to be used is selected to be surveyed and a tax data are used as auxiliary data to improve the efficiency of the estimates based on this small sample. The use of tax data in this way has allowed a significant reduction in the sample size required. The QSI is using a combination of classical sampling methodology with the use of only tax data for units where they are available. In addition to these two surveys, options to increase the use of annual tax data in a redesign of the annual Unified Enterprise Survey (UES) are presently being investigated. In this paper, we present the methodology implemented in the two sub-annual surveys and an option being considered to increase the use of tax data in the UES.
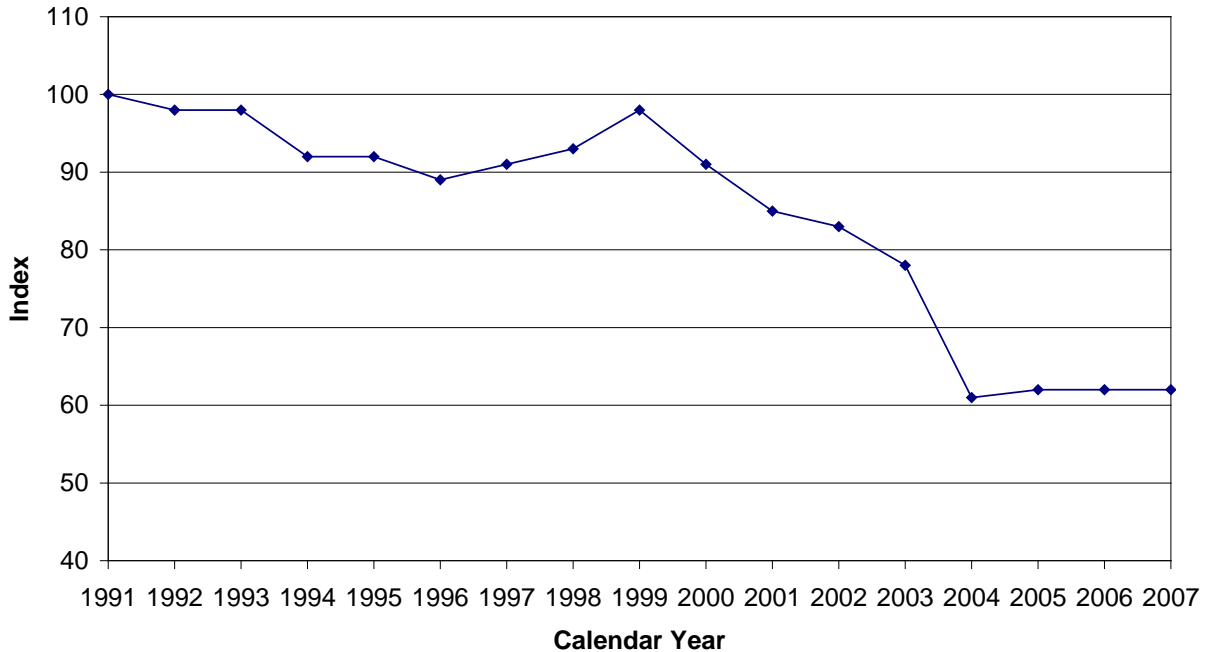
1.0 Introduction

For many years, Statistics Canada has been using administrative data for direct tabulation, frame maintenance and survey evaluation. However more recently administrative data have been increasingly used during estimation in Statistics Canada's business survey program. For the most part, Statistics Canada's business survey program has been using tax data from the Canadian Revenue Agency (CRA), the government agency responsible for collecting income tax, as a source of data to replace survey data. By doing so, Statistics Canada has been able to reduce collection costs, increase data quality and, more importantly, significantly reduce respondent burden. While there have been many challenges, some of which still exist, it is felt that the integration of tax data has been a success, in particular in reducing respondent burden. Figure 1 gives the indexed hours (1991=100) of total response burden divided by the total number of Canadian business establishments. Response burden is measured in terms of response burden hours that are a product of the frequency of each survey, the average time to complete the questionnaires, and the number of surveyed respondents. The decline in years 2000 and 2001 are most likely attributable to the introduction of a Take-None stratum, consisting of small businesses, where units are not eligible for selection. The contribution of these units is accounted for through tax data. The additional decline in 2004 is attributable to the introduction of tax replacement in both the monthly and annual business surveys. More details on the use of tax data in the business survey program are given in section 2. Given the past success of using tax data in its business survey program, Statistics Canada is now looking at ways to further increase the integration of tax data.

In this paper, we will discuss two sub-annual surveys that have significantly increased their use of tax data and preliminary plans on how annual surveys can also increase their utilization. Section 2 presents some background on the different tax data available at Statistics Canada and how they have been integrated into the business survey program. The increased use of monthly tax data by two sub-annual surveys will be described in section 3. Preliminary plans on how to expand tax data use in annual surveys will be discussed in section 4 and some summary remarks will be given in section 5.

2.0 Tax Data

The principal sources of tax data can be broken down into annual data and monthly data. This distinction is helpful as the two sources are used differently by the business survey program at Statistics Canada.

Figure 1
Index of Average Response Burden Hours
(1991=100)

2.1 Monthly Tax Data: The Goods and Service Tax (GST)

The GST is a 5% tax levied on all goods and services provided in Canada, with some exceptions. For instance there still exist some industries, such as renting, where the tax rate is 0%. All businesses with annual revenues greater than $30,000 must register for a GST account and are required to file GST remittances. The frequency of remittance depends on their annual revenue. Businesses with annual revenue greater than $6M file monthly and businesses with annual revenue between $500K and $6M file quarterly. Businesses with annual revenues between $30K and $500K are required to file annually. Quarterly and monthly filers are required to remit within 30 days of the period end, while annual filers must report within three months. Note that once a business becomes a more frequent remitter (i.e. monthly or quarterly) it remains as such even if its annual revenue falls below the threshold.

Each remittance, or transaction, consists of the business' business number, GST number, the expected filing frequency (monthly, quarterly or annually), period covered (start date and end date), total sales, input tax credit and collected GST. Each year, Statistics Canada receives approximately 9.7M transactions, covering approximately 2.6M businesses, from the CRA. In terms of counts, most of these transactions are quarterly but in terms of sales, most are monthly (see Table 1). Note that some of the quarterly remitters are actually quite small but at one point they started reporting quarterly and they are not allowed to change back to being an annual remitter.

Table 1. GST Transactions (2011 Reference Year)

	Business Counts	Transaction Counts	Sales
Monthly	8.7%	27.8%	83.5%
Quarterly	53.5%	60.7%	12.8%
Annually	37.8%	11.5%	3.7%

2

The GST data are provided to Statistics Canada by CRA seven after the end of the reference month, at which time approximately 70% of the expected transactions have been received. Expected transactions are those that are scheduled to be remitted. For example, monthly remittances are expected each and every month, while quarterly remittances are expected only at the end of each quarter. Very little data editing occurs at CRA, so once received at Statistics Canada the GST data are passed through an editing and imputation system consisting of outlier detection and imputation modules. The imputation module imputes a sales value for units that have been identified as outliers, units with missing sales figures and expected, but late, transactions. For more information on GST processing, see Brodeur and Pierre (2003).

Since monthly sales values are needed for all units, transactions that are not expected are imputed (or extrapolated) using calendarization methods (see Quenneville, Cholette and Hidiroglou, 2003). The calendarization process consists of two parts: first, sales data are modified to cover the reference month and not a different time period; second, sales data are required for transactions that are not expected. In the first part of the process, a unit's sales value is modified based on the reported period covered since even for monthly reporters, this period may not correspond exactly to the reference month. For example, a remittance may cover several days from the month before the reference month or from the month following the reference month. In such a situation, the data are interpolated (if the reporting period covers days from the month following the reference month) or extrapolated (if the reporting period covers days from the month proceeding the reference month) so that the period covered corresponds to the reference month.

The second part of the calendarization process produces monthly sales values for those units that are not expected to report. For example, if an annual reporter is expected to, and does, report in December of year *y*, a sales value is still needed for the months of January to November of year *y+1*. Calendarization uses time series methods to forecast monthly values by applying the monthly movement observed from monthly reporters to the last reported value for the unit. For more details on calendarization we refer the reader to Quenneville, Cholette and Hidiroglou, 2003). When Statistics Canada receives monthly data from CRA, approximately 70% of expected transactions are received. Although these transactions represent only 30% of the units, they account for over 61% of the total sales. The remaining 39% is accounted for by imputed or extrapolated units.

Once processed, the GST database contains monthly sales values for all units in the population and is made available to survey programs at Statistics Canada. The following month CRA provides data for the next reference month as well as any updates and additions received since the last file was provided. In order to take into account these updates and additions, which are mostly late remittances for previous reference months, the edit, imputation and calendarization processes are performed again to produce a new 'vintage' for each reference month. Thus for a particular reference month there exists files of different vintages that incorporate different amounts of reported data. Since the entire edit, imputation and calendarization systems are run, units that were imputed or extrapolated may have different values across the vintages. However, newer vintages contain more reported data and updates and are thus expected to be of higher quality. Eleven weeks after the reference month approximately 75% of the expected transactions, representing approximately 80% of the total sales, have been received.

At Statistics Canada, three mission critical surveys (the Monthly Wholesale, Retail and Manufacturers surveys) use GST data. In these surveys, a sub-sample of units is selected for tax "replacement". The units are selected from amongst the "simple" units that are selected in the sample. Simple units are those businesses that have simple structures (i.e., single industry within a geographic region) which allow a direct one-to-one link between tax data and the unit as represented on Statistics Canada's business register. By having such a link, Statistics Canada is able to assign the tax data to the appropriate industry and region combination. In contrast, complex units are those that are involved in multiple industries or have revenue generated in many regions. Tax data are not used for these units due to the difficulty of allocating it down to the industry within geography level.

When used in monthly surveys, tax data are modified through a model to account for possible conceptual differences and timeliness issues. As previously mentioned, GST is available only seven to eight weeks after the end of the reference month. Since GST data are often not available in time to meet tight production deadlines, data from month *m-1* or even *m-2* are used for reference month *m*. A sample of simple units is contacted in order to have some units that have both GST and survey data from which a model can be developed. Once obtained, the model is applied to the GST values to generate "survey" data for the units identified for tax replacement. These units are then processed as if the values were reported by the respondent. That is, the modified tax data goes through the production process (edit and imputation, review by analysts and weighted estimation) as usual.

It is recognized that from a methodological point of view, this approach is not the most efficient use of tax data as they are used for only a small number of units. In addition, by treating them as respondent data variances are more than likely underestimated. However, this approach has allowed the introduction of tax data into an existing survey program with only minor changes to the processing system and review process. Since the introduction of tax data into the Monthly Wholesale, Retail and Manufacturers surveys, two other surveys have either undergone a redesign or have been developed. Statistics Canada has seized these opportunities to significantly increase the use of GST data. More details on these two initiatives are presented in section 3.

2.2 Annual Tax Data (T1/T2)

Annual tax data are themselves divided into two sources, both available from CRA. Unincorporated businesses (sole proprietors and partners) file a T1 tax form, while incorporated (corporations) businesses file a T2 tax form. The T1 tax forms are filed by individuals to report their income to CRA for the calendar year. They have until April 30th of year y+1 to report to CRA. Minimal checks are performed at CRA and the file is then passed to Statistics Canada where more detailed editing, outlier detection and imputation takes place. Currently Statistics Canada receives only the portion of the T1 population that file electronically (approximately 80%) but does receive a file called the Assessed Record File (ARF) containing two variables (gross business income and net income) for the entire population. For estimates of variables other than gross business and net income, model based methods are used to combine the information from the electronic filers and the ARF. Estimates are provided to the subject matter areas in September of year y+1.

The other source of annual tax data, T2, covers incorporated businesses. Corporations are allowed to choose their own fiscal year but are expected to file their T2 tax information with CRA within six months of the end of their fiscal year. Because of this arrangement, CRA receives T2 tax data throughout the year and provides Statistics Canada monthly files containing T2 data for businesses that have filed in the previous month. Although CRA does some minor editing before sending the data to Statistics Canada, the majority of the statistical processing is performed by Statistics Canada. Information from the business' financial information statements (income statement and balance sheet) are provided in the General Index of Financial Information (GIFI) format. This format is a coding system that allows corporations to prepare their financial statements and report them to CRA in a standardized manner. When received by Statistics Canada, the information is passed through a series of edits to balance the data and to identify errors or outliers for correction. Any errors that cannot be corrected are flagged for imputation. Once a year, usually in September, a file is produced for the period covering April 1st of the previous year to March 31st of the current year. Prior to this release, units flagged for imputation are imputed through historical or donor imputation. For reference year 2007, the imputation rate was approximately 30%. Although this rate appears to be high, analyses have shown that the imputed data are very close to the reported data, once received. When this processing has been completed, the tax data are ready to be used by the survey programs.

Recall that a model was used to link monthly GST data to survey data to account for two possible differences: concepts and timeliness. In the case of annual tax data, a model is not necessary because data from CRA are available in time and a Chart of Accounts (COA), based on standard accounting principles, has been developed to ensure coherence between survey and tax based concepts. For more information on the COA see Vinette (2005). Thus, if a survey is COA compliant then variables asked for

on the questionnaire have a conceptually equivalent field (or sum of fields) reported on the tax form. Because of this equivalence, annual data are used in essentially the same way as GST data except no model to account for conceptual differences or timeliness issues is necessary. Therefore, a sub-sample of simple units does not go the field for collection and their financial variables are taken directly from tax data. Note that since only financial variables are available from tax, no information on commodities, or non-financial variables, is obtained for these tax replaced units.

Within the annual business survey programs there are currently three slightly different implementations of using tax data, largely the result of when tax data were introduced into the surveys. In the first approach, a sample of simple units is selected for tax replacement. Those simple units not selected for tax replacement, and selected complex units, are sent a questionnaire covering financial and non-financial variables. Financial information for 'tax-replaced' units is obtained directly from tax data. Non-financial information is typically imputed using data from those simple units who receive a questionnaire. Once the necessary information is obtained for all units, sample weights are used to obtain weighted estimates. The second approach is essentially the same as the first one except that simple units not selected for tax replacement receive a shorter questionnaire covering only non-financial variables. Financial information for these units comes from tax data. Complex units continue to receive a questionnaire covering both financial and non-financial variables. The third approach is also similar to the first approach except that financial information is obtained from tax data for all non-sampled simple units in the population producing a 'pseudo-census'. Weighted estimates of non-financial variables are based on responses from sampled simple units and are calibrated to the financial totals coming from this pseudo-census. These estimates are then added to those obtained from the complex units.

Statistics Canada is currently planning a major redesign of its annual business surveys which make up the Unified Enterprise Survey (UES). In the late 1990s the UES was developed and consolidated all annual business surveys into an integrated framework. It currently consists of over 60 surveys, the majority of which are using tax data in one of the three approaches recently described. For more information on the UES, see Brodeur et al. (2006). A goal of the redesign is to maximize the use of tax data and to develop a single approach that would be applicable for all surveys. More details on a sample design being investigated for the UES redesign are given in section 4.

3.0 Increased Tax Use in Sub-Annual Surveys

Two sub-annual surveys have been recently redesigned or developed and have made more extensive use of monthly GST data than the three previously mentioned surveys (the Monthly Wholesale, Retail and Manufacturers surveys). The first survey, the Monthly Food Services Survey (MFSS), has recently been redesigned and and is using the calibration approach, with tax data as auxiliary data, to reduce the sample size necessary to achieve targeted levels of quality. The second survey, the Quarterly Services Indicators Program (QSI), was developed with the intention of using GST data to the maximum by using them for all simple units but, unlike the other monthly surveys, does not apply a model. More details on these surveys are given in the following sub-sections.

3.1 Monthly Survey of Food Services and Drinking Places (MFSS)

The MFSS produces estimates of the value of sales and the number of locations of restaurants, caterers and drinking places as defined by the North American Industry Classification System (NAICS). As the survey used little tax data in the past, one of the goals of the redesign project was to increase their use. As noted earlier, the first uses of GST data were done in such a way as to not affect on-going production systems and processes. Because the MFSS was undergoing a major redesign, it was decided from the outset that an increased use of tax data would be imbedded into the design from the start.

The MFSS sample design combines elements of a typical design used for business surveys at Statistics Canada with a strategy that takes full advantage of tax data. In both designs, the first step in stratification is to stratify by NAICS code and geography (provinces and territories). Within each NAICSxGEO group the smallest units that represent at most 10% of the total revenue of the group are placed in a Take-None (TN) stratum and are ineligible for sampling. How the contribution of these units is estimated through the

use of tax data will be explained shortly. After the TN units are identified, Take-All (TA) units are then identified. These units include large units that are identified using the Lavallée-Hidiroglou algorithm (Lavallée and Hidiroglou, 1988) or by subject matter experts and represent themselves. Once the TN and TA units are identified, the sample design depends on the number of simple units remaining. In those NAICSxGEO groups where there are only a few simple units, a traditional survey design is used. That is, units are selected and questionnaires are mailed out. Once returned, the data are captured, processed and weighted estimates are produced.

However, in those NAICSxGEO groups where there are a significant number of simple units a further stratification is done according to whether or not they are complex. Complex units in these groups are handled using the same traditional survey design described above (i.e., a portion of units are selected and are mailed a questionnaire). In the strata of simple units, a small sample is drawn and these units are sent a questionnaire. During estimation a ratio estimator is used in these strata with the tax data being used as auxiliary data. As one would expect, the tax data is highly correlated to the variable of interest (monthly sales) which has allowed a significant reduction in the size of the sample necessary to maintain an appropriate level of precision. By including the TN units in the calculation of the population totals used in the ratio estimator, their contribution to the survey estimate is automatically included.

This survey design has allowed Statistics Canada to significantly increase the amount of tax data used by the MFSS and to significantly reduce respondent burden. In the past, the MFSS used an approach similar to that employed by the Monthly Wholesale, Retail and Manufacturers surveys and used GST data for approximately 13% of its units (representing approximately 20% of total sales). With the new design GST data is used as auxiliary data for 85% (approximately 25,000 units) of its units and represents 54% of totals sales.

While the new MFSS has been considered a success, some challenges have arisen as a result of the increased use of tax data. A major challenge is the stability of the ratio estimator based on GST data. Since many of the simple units used to calculate the auxiliary data totals are annual or quarterly reporters, a large portion of the GST data have been extrapolated through the use of calendarization methods. As with any forecast there is some variability introduced. This variability, combined with a very unstable industry, shows up as instability in the ratio estimator. An additional challenge arises in the revised figures that are published each month. In addition to the estimates for the reference month, the MFSS publishes revised estimates for the two previous months. These revised figures included any late responses or updates received from survey respondents and late GST remittances to CRA. However, in addition to the late GST remittances the revised figures include updated imputed and extrapolated values coming from the GST processing. Work is currently being done to investigate how much of the revisions being seen is a result of updated imputed or extrapolated values.
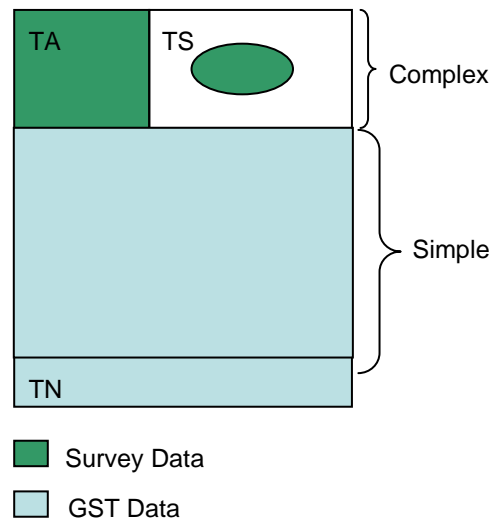
3.2 Quarterly Services Indicator (QSI) Project

As in most countries, the service sector represents an increasingly large portion of the economy; however, our knowledge of this sector remains somewhat incomplete. While Statistics Canada has a strong program of annual Services surveys, there is a lack of quality data at the sub-annual level. Until now Statistics Canada has relied heavily on employment data to produce sub-annual measures of output for the services sector. Given that many services industries are dominated by businesses that tend to be small non-employers, the use of employment data as a proxy for output data poses a risk of bias. In response, Statistics Canada started development of the QSI which relies heavily on the use of GST data to produce quarterly movements of sales for 41 service industries. In order to be relevant these indicators are to be published 90 days after the end of the reference quarter at the four digit NAICS level for all provinces and at the national level. In addition, revised estimates will be available 90 days later (i.e., 180 days after the end of the reference quarter).

Early in the development of the QSI it was realized that the use of GST data for complex units would be problematic.   Although allocated GST data existed, the allocation methodology was still under development and its resulting quality was unknown. On the other hand GST data for simple units, which did not require allocation, was a rich source of timely data. The decision was made to use traditional

survey methods for the complex units and GST data only for the simples. Estimates for the complex and simple units would be calculated separately and then combined to produce level and quarter-to-quarter trend estimates. Given that no model would be used to adjust the GST data for potential conceptual differences, the resulting estimates of level were to be benchmarked to those obtained from the corresponding annual surveys. The QSI design is illustrated in Figure 2.

Figure 2
Quarterly Services Indicators Sample Design

While the survey of complex units has experienced some challenges (low response rates, lack of historical data for edit and imputation purposes, etc.), we focus the discussion on some of the challenges faced as a result of the significant use of GST data for simple units. One issue faced by the surveys using GST data is the ability of tax data to accurately identify births, and more importantly, deaths in the population. The QSI is more susceptible to deaths in the population than the MFSS because it has been decided to accept a signal of a death from tax data only, while the MFSS waits for a signal from the business register. As explained earlier, calendarization techniques are used to produce monthly sales figures for all units in the population. For annual reporters, this amounts to forecasting sales for 12 months until their next annual remittance. As there is no contact with the unit, there is no way to know if it is still active or not. In fact since not all businesses remit on time, Statistics Canada continues to impute a monthly GST sales value for six months after the expected remittance month (for annual reporters). If information has still not been received after these six months, it is considered as inactive and its monthly sales value is set to zero for all months since the last remittance. This approach of deathing a unit is more liberal than that of the business register which waits for a confirmation of cessation of business before making the unit inactive. If the MFSS does not receive tax data for a unit that is still considered active on the business register, a value is imputed for the reference month. On the other hand, if the QSI does not receive a GST value then it is considered as inactive and a sales value of 0 is assumed. It does happen that after a unit is declared dead by the tax data approach, it does remit at which time it is reactivated causing additional noise in the estimates.

This inability to know the status of a business causes two problems: overestimation of sales during the period that a positive value has been imputed when in fact the unit is inactive and a possible break in the series when the unit is finally deathed. Given that the QSI was conceived to estimate quarterly movements, one could argue that this overestimation of sales would cancel out and the effect on the estimates of movement would be only minor. The second problem of deathing a unit can severely affect the revised estimates as the initial estimate contains positive sales values that no longer contribute to the revised estimates. Alternate sources of information to accurately identify inactive units in a more timely fashion are being investigated.

7

Another issue facing the QSI is the different vintages of GST data for the same reference month. A major concern during the QSI development was the change observed between the initial estimates of movement and the revised estimates. A large portion of this change was attributed to the 'reprocessing' of data after updates and additional remittances have been received. Given that all processes are redone, some units that were initially imputed or extrapolated are once again imputed or extrapolated. One could argue that these new values are 'better' in the sense that they are closer to the true values once reported; however, preliminary analyses indicate that they are not 'better' than the initially imputed or extrapolated ones. By re-imputing or re-extrapolating these units, an additional source of variability has been added that is not necessarily improving the quality of the estimates. One option being considered is to freeze the imputed or extrapolated values until a reported value is received or a signal from another source is obtained that indicates that the unit is no longer active. The effect of this option is currently being investigated.

4.0 Increased Tax Use in Annual Surveys

The Unified Enterprise Survey (UES) was developed over ten years ago. In the first year only seven industries were in scope, but since then it has evolved such that it now encompasses more than sixty different industries. As surveys were integrated into the UES, many changes were made to the methodology and processing infrastructure to accommodate program needs. Following ten years of evolution, Statistics Canada believes that it is now time to take a step back and rethink the entire UES methodology and processing framework. This exercise would be an excellent opportunity to ensure that the redesigned UES takes full advantage of tax data.

Unlike the sub-annual indicator surveys discussed in sections 2 and 3, the UES is interested in more than just annual sales figures. In addition to annual sales, the UES questionnaire asks for detailed revenue and expense items and if applicable, data on commodities produced. The need for detailed financial and commodity data pose particular challenges for the use of tax data as this information is not always available. Although the COA provides a direct mapping from tax data to survey variables, not all fields on the tax return are mandatory. In fact only eight fields are mandatory, essentially corresponding to the financial totals (e.g., Total Revenue, Total Expenses, etc.). While some units do report their detailed financial data, not all of them do. For those units that do not report their financial details, values are derived using the mandatory totals and a distribution obtained from those units that have reported the details. Analyses have shown that the mandatory tax fields compare very well with reported survey data. Thus the imputation strategy uses these 'totals' as anchors and ratios are applied to obtain the details. While this method produces quality estimates at an aggregated level, analyses have shown that, as expected, the quality of the micro-data is not as good.

Commodity data are not available from tax data and efforts to use statistical models to relate financial data to commodities have proven to be futile. That is why in all three currently used approaches some simple units are sent questionnaires (either full questionnaires covering both financial and commodity variables or short questionnaires that cover commodity information only). Weighted commodity estimates, calibrated to financial totals or not, are then produced. We note that the commodity estimates are domain estimates since stratification is based on annual revenues. In the future, if a census of financial information comes from tax data (for the businesses that are simple in structure) but estimates for commodities continue to come from a probability sample, then the sample should be designed to produce quality commodity estimates. As of now a design variable that is highly correlated to commodities has not been found.

Another concern that must be taken into account if significantly more tax data are to be used is the lack of survey feedback that would be received if tax data were to be used for all simple units. A significant amount of information about a unit's status (active or inactive) and its industrial activity is obtained during survey collection. This information is fed back to Statistics Canada's business register which acts as a frame for essentially all of Statistics Canada's business surveys. If the use of tax data use is to be increased and this survey feedback is lost, the status and industrial activity of a large number of units could become out-of-date very quickly.
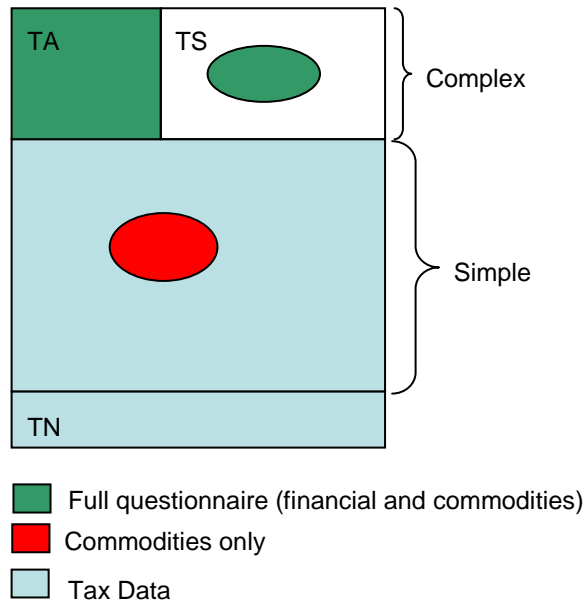
One model being considered for the UES redesign is similar to the third approach discussed in section 2.2. Due to the difficulties of obtaining tax data at the appropriate level for complex units, these units would continue to be surveyed. Tax data would be used for financial variables for all simple units in the population. To ensure the quality of the resulting estimates, some work needs to be done on improving the financial details coming from tax data. Estimates of commodities for simple units continue to pose a challenge. One option would be to survey a small number of simple units and to ask questions only on commodities. The quality of the resulting commodity estimates could be improved by calibrating to financial totals obtained from the pseudo census of tax records. The status and industrial activity of these simple units could be obtained during the contact required to obtain commodity data and used to update the business register. The size of this sample of simple units would have to take into account the trade off between respondent burden, the quality of commodity estimates and up-to-date information on the business register. This possible design is illustrated in Figure 3 on the following page.

5.0 Summary and Future Challenges

The use of tax data in the monthly business statistics program at Statistics Canada has been significantly increased as demonstrated in the MFSS and the QSI. While many challenges have been overcome, there are still some more that are a direct result of using significant amounts of tax data. While the use of tax data in the annual program has not kept pace with the monthly programs, a major objective of the upcoming redesign is to significantly increase their use. Some preliminary plans have been investigated and research has begun on solving the challenges mentioned in this paper.

While most of the recent advances have been in the direction of using tax data for more units, Statistics Canada is also looking into ways of broadening and deepening its use of tax data. A large amount of additional information is available from tax data in the form of schedules. These schedules provide a wealth of information on certain aspects of businesses that file them. Statistics Canada has begun investigating the quality of the data obtained from these schedules to establish if they can be used in its statistical programs. In addition to the new information coming from these schedules, work continues on developing methods to increase the number of variables that can be taken directly, or derived, from the tax data. Finally, Statistics Canada is looking to expand the use of tax data to programs outside its business survey program such as Balance of Payments, Capital Expenditures and the System of National Accounts. While Statistics Canada is very interested in increasing its use of tax data in its business survey program, an important factor that needs to be considered before making the decision to use tax data for all simple units is whether or not Statistics Canada is willing to accept a possible break in data series that may occur if such an approach is implemented.

Figure 3
Possible Unified Enterprise Survey Design

References

Brodeur, M., Koumanakos, P., Leduc, J., Rancourt, E. and Wilson, K. (2006). The Integrated Approach to Economic Surveys in Canada, Statistics Canada Publication 68-514-XIE, Statistics Canada

Brodeur, M. and Pierre, L. (2003). Use of Tax Data: An Application of Goods and Services Tax (GST) Data, Proceeding of Statistics Canada Symposium 2003, Statistics Canada.

Lavallée, P. and Hidiroglou, M. (1988). On the Stratification of Skewed Populations, Survey Methodology, 14, 33-43.

Quenneville, B., Cholette, P. and Hidiroglou, M. (2003), Estimating Calendar Month Values from Data with Various Reporting Frequencies, Proceedings of the Business and Economic Section of the American Statistical Association.

Vinette, L. (2005). Progress in Collection Business Data, Proceeding of Statistics Canada Symposium 2005, Statistics Canada.

Yung, W., Rancourt, E., Hidiroglou, M. (2007). Administrative Data in Statistics Canada's Business Surveys: The Present and the Future. Seminar on Registers in Statistics - methodology and quality, Helsinki, 2007.