

USE OF TAX DATA IN THE PRODUCTION OF PROVINCIAL ECONOMIC STATISTICS

Peter D. Bissett

Statistics Canada

Abstract

In late 1996, Statistics Canada embarked on a major project to improve the quality of its provincial economic statistics (PIPES). The project goal is to implement a full-scale system of annual economic statistics by province, consisting of business and household surveys combined with data derived from tax and other administrative data sources.

This paper provides the background to PIPES, reviews the various tax files being exploited and describes the new electronic financial statement reporting requirements for Canadian corporations. It also describes how Statistics Canada converts these administrative files into statistical information and provides an overview of the role of administrative data in the production of provincial level economic statistics.

Introduction

In the middle 1990's the Canadian Federal Government promised to replace the existing Goods and Services Tax (GST) " with a system that generates equivalent revenues, is fairer to consumers and to small business, minimizes disruption to small business, and promotes federal-provincial fiscal co-operation and harmonization". In April 1997, a new "Harmonized Sales Tax" (HST) was introduced in the provinces of Nova Scotia, New Brunswick, and Newfoundland and Labrador. The HST replaced the GST and the three different provincial sales taxes with a single value-added tax in the three participating provinces.

The HST levies tax on the value added at each stage of the production process of both goods and services. The revenues generated by the tax are collected by Revenue Canada (Canada's central tax agency) and "pooled" in a central fund. A HST Revenue Allocation Formula divides these revenues among the participating federal and three provincial governments.

Statistics Canada was asked to provide the statistical information required for the revenue allocation formula. Thus was born the Project to Improve Provincial Economic Statistics (PIPES). This request provides a major challenge to Statistics Canada on a number of fronts, including the challenge of minimizing respondent burden. A tremendous amount of detailed information is required to distribute accurately the pooled tax revenues among the four participating governments. If each business filing HST were obliged to collect, maintain and report on all the data necessary to allocate the pooled revenues properly a significant burden would result. It is obvious that this situation is not acceptable for several reasons, including its undermining of two major benefits which the HST was to promise: to lower compliance costs and increase simplicity in the federal and provincial tax systems.

The paradox faced by Statistics Canada is that it must provide large volumes of detailed provincial economic data (the majority of which traditionally is obtained via surveys of Canadian businesses) without undue increase in respondent burden. Moreover, the prevailing view at Statistics Canada is the perception that the "goodwill" of small business is already stressed by the current business survey

The result has been the development of the mantra to maximize the use of administrative data sources. The remainder of this paper explores the challenges and opportunities associated with PIPES and the heavy reliance on administrative data sources or the "tax to the max" position.

The Landscape

The requirements of PIPES stipulate that Statistics Canada produce statistics for each of the ten provinces and three territories with roughly equal quality. These statistics are required for the full suite of industries as described by the North American Industrial Classification System (NAICS). The level of quality across industries may vary based on a

number of factors such as the importance of a given industry and its relative level of economic activity for a given geography. Finally, these statistical tables are required annually.

While Statistics Canada has produced provincial level statistics for many years, the existing programs fall well short of the data quality requirements of the HST. In order to measure the final sales of goods and services accurately, on an annual basis by province in sufficient detail, Statistics Canada must produce a detailed set of accounting entries – including Provincial input/output tables. With approximately 300 industries, 700 commodity groups and ten provinces and three territories there is the potential of tables containing upwards of 2 million cells.

There are approximately 1.7 million businesses with which to populate the provincial I/O tables. The vast majority of these businesses (>99.5%) operate in a single industry within a single province. The use of administrative records to obtain economic data for these businesses is possible.

Problematic are the relatively few "complex" businesses. Complex businesses are defined for our purposes as those businesses operating in more than one industry, and/or more than one province, and/or who are comprised of more than one legal entity. Complex businesses comprise less than 1% of the Canadian business universe yet account for more than 50% of the gross domestic product.

For the non-complex businesses, one can compile the tax data and associate those data directly with the single province and industry of the business. For the complex businesses, life is more challenging since one must find a way to break down such businesses into separate industry and province parts. The tax data are not of much help in accomplishing the latter task. It is, therefore, necessary to continue to survey portions of the Canadian Business universe as a means to determine the provincial and industrial components of complex businesses. Surveys also collect information not available from administrative sources such as class of customer and commodity detail.

Administrative Files

Administrative data uses have a long history at Statistics Canada. Since the early sixties administrative data have progressed from being simply a "tool" to help reduce response burden (primarily on small business) to becoming an integral data source for many important statistical programs. Much of the successful integration of these data sets is a direct result of the excellent cooperative relationship existing between Statistics Canada and Revenue Canada. As per the Canadian Statistics Act, the information flow between the two departments is one way. The Statistics Act permits Statistics Canada to access and use the tax data sets for statistical purposes but prohibits Statistics Canada from sharing survey data with Revenue Canada or any other such agencies. Statistics Canada and Revenue Canada collaborate in the assignment of industrial codes (NAICS). Revenue Canada provides information on the business activity; Statistics Canada assigns a NAICS code, which it shares with Revenue Canada. A number of institutional arrangements exist including annual meetings of deputy heads, regular liaison committee meetings and well defined focal points within each department. These various communication entities permit the organized flow of data, allow each department to understand the various needs and help focus on common goals such as the minimization of paper burden on Canadian business and citizens.

Revenue Canada makes available to Statistics Canada a wide range of administrative files:

Unincorporated Returns (T1)

Approximately twenty million T1 income tax returns are filed annually by individuals. Presently more than three million of these T1 tax filers report business, professional, farming, fishing, commission or rental income to Revenue Canada. More than one million of these use the electronic version of the Standardized Financial Data (SFD) form, designed jointly by Statistics Canada and Revenue Canada to report their financial information. The remaining two million unincorporated tax filers continue to use paper filing to report this type of information – though the number of businesses in this category is dropping every year. Each year, Statistics Canada draws a sample of electronic and paper documents of T1 tax filers whom report business income and generates industry estimates.

Incorporated Business Returns (T2)

Annually corporations reporting income from any source file a T2 return. These corporations are required to submit their tax return within six months of their fiscal year end. Each year approximately 1.2 million T2s are filed. Every year, Statistics Canada receives a universe file containing information such as "name and address" of each corporation and a basic set of financial variables - assets, equity, sales, profits and taxable income (with provincial breakout).

Generalized Index of Financial Information (GIFI)

The Generalized Index of Financial Information (GIFI) is a standardized financial statement. GIFI is being introduced to facilitate the electronic filing of Canadian corporate tax returns. The GIFI defines a standard set of Income Statement and Balance Sheet items that corporations are required to file when reporting the results of their business activities to Revenue Canada. The data are captured electronically and made available to Statistics Canada annually.

Goods and Services Tax (GST) and Harmonized Sales Tax (HST) Files

Statistics Canada has access to the Revenue Canada GST and HST remittance files. The files contain each registered business' activity code, accurate contact information, total sales of GST/HST liable goods; the GST/HST value collected, and the GST/HST on inputs purchased by each business (and thus for which an input tax refund claim is filed). Every business in Canada earning \$30,000 or more in annual sales is required to register for a Goods and Services Tax or Harmonized Sales Tax account with Revenue Canada. These businesses are required to submit records of the tax collected and to claim input tax credits. Smaller firms may also choose to establish an account, in order to claim credits. There are roughly 2 million active GST/HST accounts on the Revenue Canada file. Statistics Canada receives GST/HST files on a monthly basis.

Payroll deduction

All employers in Canada are legally required to establish payroll deduction accounts. There are approximately 1 million active accounts. Payroll Deduction (PD) Accounts are established as a means to remit to Revenue Canada employer and employee contributions towards Canada/Quebec Pension Plan and Employment Insurance, as well as source deductions for income tax.

The file contains information on remittances, employment, and total payroll for the firm. The two latter items were added at Statistics Canada's request to provide data for the Survey of Employment, Payroll and Hours (SEPH).

Customs Documents - Import and Export Shipments

The International Trade program within Statistics Canada has access to a continuous flow of Customs administrative data reflecting all Canadian import and export shipments. These data are compiled monthly to derive value and volume statistics by country, commodity and province of clearance/origin. Canada and the United States of America (USA) have an arrangement whereby they exchange their import data, providing a scrutinized source of export data for each country.

Tax Data – Generalized Index of Financial Information (GIFI)

As mentioned earlier, Statistics Canada faces the paradox of having to provide a wide range of detailed data for each of the provinces and territories without introducing an undue increase in response burden on Canadian businesses and citizens. GIFI did not exist when PIPES started in 1996, although the idea was gaining momentum at Revenue Canada. In a bold move Statistics Canada management decided at that time to cast its lot with GIFI as a way of pursuing the goals of PIPES without raising survey response burden excessively. While GIFI was by no means a sure bet, Statistics Canada shared development costs and worked closely with Revenue Canada to nurture the project along. At last, the gamble is beginning to pay off.

Generalized Index of Financial Information (GIFI) - Background

GIFI provides an exciting data source containing detailed revenue, expense, profit, asset, liability and net worth figures for all Canadian companies big and small. Using this unique data source, Statistics Canada will produce precise

statistics at the sub-provincial level, by 6-digit NAICS industry and by size group. We are moving from an era where we depend on surveys with small samples, focused on the large companies and geared towards producing aggregate macro data, to one where we depend on census data from administrative sources to produce detailed regional, industrial and size-group level estimates.

Critical to maximizing the use of GIFI is the determination of how the GIFI data meet the content and definitional requirements of the statistical program. For example, analysis of the ability of the GIFI to provide the detail required to calculate "value added" is currently underway. Extensive work is proceeding on "mapping" GIFI variables to the many survey variables. Accountants, working with subject matter experts, create tables showing which variables on the questionnaires are conceptually the same as items within the GIFI template. The tables also illustrate the nature of the relationship. The conceptual relationship may be one to one, one to many, many to one or many to many. There may be no relation as GIFI contains only financial statement data. Questionnaire detail such as commodity or class of customer will not exist within GIFI.

The result of this work has already led to both removing questions and to creating short questionnaires (questionnaires which do not contain any questions on financial details). In each case, GIFI data will be used in lieu of survey data. For some industries there will be no survey for reference year 1998. Instead, the results of the 1997 questionnaires, in combination with the 1998 GIFI data, will be used to generate statistics using modeling techniques.

GIFI is the result of an effort to permit electronic reporting of corporate tax returns. An obstacle in realizing this goal revolved around the requirement that businesses report their financial statement with their tax submission. In Canada, as in most countries, companies have wide leeway in how they lay out their financial statements. Revenue Canada had to find a way to standardize accounting statements that would not alienate the accounting community. Drawing on a diverse sample of financial statements, they studied the variety of line items reported. Using this sample they derived an "index" of line items and assigned each a unique code. By requesting businesses to assign one of the unique codes to each line item of their financial statement, the businesses are not required to change their format and Revenue Canada is able to introduce electronic filing. The GIFI concept has received wide acceptance among the business community and is now a fait accompli.

It is important to recognize that Revenue Canada has developed the GIFI structure in conjunction with major accounting associations and other major stakeholders such as Statistics Canada. Working closely with accounting software manufacturers has resulted in several accounting software packages being available which support the assignment of GIFI codes to financial statements. Once the information is "coded" the software will create a GIFI formatted tax return at the push of a button.

GIFI is an extensive list of financial statement items where each item has a unique standard code (e.g., cash is 1001). The structure of GIFI facilitates the selection of items usually reported on a business' financial statements but in a standardized format. The GIFI information has to balance. To verify the accuracy of the information Revenue Canada uses the following rules:

- total assets = total liabilities + total shareholder equity
- net non-farming income = total revenue – total expenses
- net farm income = total farm revenue – total farm expenses

While GIFI provides an extensive list of financial items not all businesses report the detail. GIFI is structured hierarchically with detailed items grouped into "blocks", "blocks" are grouped into "sections", and "sections" are grouped into "mandatory" items. The mandatory items are those fields required to verify that the GIFI return balances (see equations above). Each business filing GIFI has the option of reporting a combination of detailed items, block level items, section level items and/or mandatory level items. Our experience in analyzing the first 300,000 1998 GIFI returns indicates that on average 46 items of the various levels are being reported per GIFI return.

GIFI – Statistical Information

The Tax Data Division (TAX) has the responsibility of coordinating the receipt of the GIFI data, converting it to statistical information, and delivering the result (according to pre-approved specifications) to the PIPES internal

clients.

For 1998, the GIFI data passes through three phases. First, the "raw" data are received from Revenue Canada and transferred from tape to a local server. A number of consistency edits are applied such as checking for duplicate records and completeness of data items. When these checks are completed, the "raw" data are made accessible to all authorized users.

The second processing stage is referred to as the "cleaning" stage. Each micro record is examined against a basic set of checks. These include verifying that the mandatory fields are completed, the record balances, certain base variables are derived, etc. Again once this stage is completed authorized users are notified and access provided.

The final process consists of a series of edits and imputations resulting in the production of a "statistical" file. Process timelines for various economic programs including PIPES is such that a universe of GIFI files are required before receiving all of the GIFI returns from Revenue Canada. The final GIFI files are not expected from Revenue Canada until the end of the first quarter of reference year y+2. As such, it is necessary that TAX impute the data for those records still being "cleaned" or not yet received.

Partial imputation is performed at the "section" level. It is performed for returns with missing section or unbalanced sections (sections include the balance sheet and income statement). Total imputation occurs when the record is missing or does not balance within a pre-specified tolerance. Both imputations are performed using Statistics Canada's Generalized Edit and Imputation System (GEIS) and is based on donor imputation.

It is important to state that Statistics Canada places a high priority on maintaining the confidentiality and security of all its data. This is especially so for administrative data as any breach of security would have major negative consequences not only for Statistics Canada but, for Revenue Canada and the government of Canada. Thus, there are strict procedures limiting access too only those within Statistics Canada who are authorized to use the data.

Use of Administrative Data

The reliance on administrative data sources has increased considerably over the years. Initially the use of administrative data was seen as a way to reduce response burden imposed on the small businesses. However, administrative data have also become vital to the very maintenance of important statistical programs. Apart from providing the basis of and signals for the Business Register (it is used to detect deaths, births and restructuring of businesses), administrative data are extensively used in the Agriculture and Labour Statistics programs. As well, they are a major input source for analytical projects such as the production of small business profiles of income and expenses and selected balance sheet items. The PIPES initiative is defining an environment in which our reliance on the use of administrative data, and in particular, tax data, will increase even further.

Unified Enterprise Statistics Program (UESP) – A New Approach

The UESP is a new philosophy of business statistics production that emphasizes the enterprise over the establishment. The primary incentive for this change in philosophy is the need to better manage our respondent relations. This enterprise-centric approach encourages common, harmonized survey production systems for questionnaires, sampling, collection and post-collection processing and stresses the importance of using existing administrative data sources rather than surveys whenever possible. There are four fundamental principles of this new approach.

These are:

1. The unit of statistical observation and primary contact is the enterprise – in contrast to the establishment, the legal entity and the various other units currently in use.
2. Administrative data are used in preference to survey vehicles. Corporation and personal income tax returns, GST compliance forms, payroll deduction statements, customs forms and other administrative sources will be used to the fullest extent possible. Maximum use will be made of the Revenue Canada Business Number and generally, direct burden will only be increased if there is no satisfactory administrative alternative.
3. There is only one survey frame – the Business Register.

4. A central contact management system is built which measures response burden by enterprise.

As point 2 indicates, administrative data are being used wherever possible in direct replacement of survey data. If an enterprise has only one legal entity and a set of production units in one industry and one province, the data may be used as an enterprise-level benchmark total and as a source of information about production.

If the enterprise has only one legal entity and its set of production units covers more than one industry or more than one province, the administrative data can be used for enterprise level statistics. Using the data for production statistics, however, is only possible if the profile of the enterprise is known and appropriate allocation variables have been provided.

If the enterprise is still more complex (i.e. with more than one legal entity and/or both covering more than one industry and more than one province) the administrative data cannot be used directly for either the enterprise statistics or the production statistics unless a profile of the complex elements of the enterprise exists and when consolidation and allocation variables are provided. In summary, regardless of the complexity of an enterprise it is possible to use the administrative data as long as a profile with appropriate allocation data is available.

Administrative data is used as a basis for sample stratification and sample size allocation. For example, many economic surveys use the gross business income (GBI) as a stratification variable. At Statistics Canada, GST data, PD accounts data, GIFI data or data modeling techniques can all be used to populate this variable.

The introduction of the Business Number (see next section for more details on the Business Number) permits greater use of administrative data for editing and imputation of survey data. This is a result of the improved linkage between tax units and survey units. As a minimum, the GIFI data are used to identify which survey units to use as donors for other survey non-response units. In other situations, the administrative sources provide data with which to edit and impute data across sources. Finally, it is possible to construct a complete set of administrative variables for the entire universe of enterprises.

Expectations are greatest for use of administrative data at the estimation stage. The benefits expected are reduced respondent burden, reduced survey cost and increase in the reliability of survey estimates. Building on our experience using administrative data sources for modeling techniques, Statistics Canada expects to be able to project sample results across the entire population by means of the strategies embodied in the UESP. Thus, GST and GIFI data provide a detailed and comprehensive picture of the revenues, expenditures, assets and liabilities of simple enterprises. Additional variables that are required under PIPES – commodities purchased and sold, origin and destination of shipments, class of customer information and so on – are being measured with small sample surveys and then associated by models with the administrative data.

Generally speaking, the risks associated with any administrative data source are difficult to measure. Responsibilities for the content and implementation plan rests with another Federal Government Department and, as such, are not under Statistics Canada control. In the past few years, however, the climate has changed to the point where Statistics Canada now is a significant partner in the design and implementations of broadly based government initiatives. During the transition process to GIFI, careful quality evaluations of the administrative sources are being performed. Statistics Canada is monitoring and actively encouraging adherence to time schedules.

Issues of interest

Introduction of the Business Number

Several issues including timeliness, linkages, data elements and cost traditionally have hampered the use of administrative data.

The fact that companies have six months after the end of their fiscal year to file their income tax form introduces a significant delay in the process of acquiring tax data. In addition, the traditional process of transforming tax data into statistical information has been a lengthy process of data capture, editing and analysis of a sample of non-standardized businesses' financial statements. The introduction of an electronic standardized GIFI schedule should reduce the time

lags. Electronic reporting will decrease processing (eventually no more manual data capture) times. Standardized concepts will facilitate the process of converting the data into statistical information.

Presently, the financial statement data are received in a variety of free form formats. The transformation of these data to a standard set of statistical items is fraught with problems of definition, interpretation and completeness. In the past, one of the most problematic issues was the lack of linkages between administrative units (i.e. corporations filing tax returns), the units of interest to Statistics Canada (i.e. business entities surveyed), and other various administrative data sources. The variety of registration and numbering systems for administrative programs resulted in data sets that often were not easily linked, thus sub-optimizing the use of the data for statistical purposes. However, major improvements have been realized on this front in recent years.

In its February 1992 budget, the Canadian Government announced an initiative to establish a "single business registration number" as a measure for "making it easier to deal with the Federal Government". The budget papers stated:

"At present, different departments require as many as six different registration numbers from Canadian businesses. This means more work for businesses, more cost for government and, inevitably, poorer service. The government is committed to making the necessary changes to arrive at a single registration number in cases where this would be advantageous to the business concerned. This will mean less paperwork for business, greater efficiency and responsiveness by government, better economic statistics and a more effective system of revenue collection."

The proposal of a single number which businesses could use when dealing with the Federal Government was accepted and is now implemented as the "Business Number (BN)". Revenue Canada issues the BN to anyone requesting Payroll Deduction, GST, Importer-Exporter or T2 accounts. Revenue Canada has completed converting all pre-BN business accounts to the BN system. Now that the BN system is fully implemented it is simpler to associate or link different tax data sets and survey records for a particular business (legal entity).

Furthermore, Statistics Canada has an agreement with Revenue Canada whereby Statistics Canada codes all new BN accounts with a North American Industrial Classification System (NAICS). This is yet one more step in facilitating the linking of administrative data and survey data and in the conversion of tax data into reliable statistical information. The two agencies have a very effective partnership.

Legal Entity – Statistical Enterprise Complex Enterprises

There are about 2 million business enterprises in Canada with revenues greater than \$30,000. Approximately one million enterprises are incorporated and of these, approximately 8,000 enterprises are "complex" in the sense of operating in more than one province or more than one industry or both. The implementation of NAICS the effect of slightly increasing the number of "complex" enterprises. This is due to NAICS having more detailed industries, especially in service industries and therefore, there is a greater chance that multi-establishment enterprises will cross industry boundaries under NAICS.

An essential challenge of the UESP is to collect reliable, complete and consistent provincial information on sales, expenses, product, customer and location of establishments associated with these 8,000 multi-province, multi-industry enterprises without imposing on them an intolerable response burden.

Confidentiality and Data Release Issues

The production of comprehensive provincial economic accounts represents a "new era" in terms of the amount of statistical detail available. Input/output tables of relatively comparable quality will be produced annually for the 10 provinces and 3 territories. The excitement associated with the knowledge that these data will be produced is tempered by the fact that with several small provinces (small populations) and territories much of the data will not be released due to confidentiality constraints.

Moreover, current legislation restricts the sharing of administrative data beyond the bounds of Statistics Canada. This is problematic from the point of view of response burden minimization. In the past, Statistics Canada has actively

shared its survey data with provincial statistical agencies. In situations where administrative data are used and Statistics Canada is restricted in sharing these data the partners may feel obliged to conduct surveys themselves.

Despite these constraints, the value of the data for numerous and wide ranging applications within Statistics Canada remains very high. Statistics Canada has also, in the first year of PIPES, produced and released economic data for a range of industries not covered for many years or at all. These include aquaculture, construction, taxi and limousines, lessors of real estate, couriers, and restaurants and taverns. These industries represent approximately 15% of the Canadian economic activity and 20% of all Canadian businesses by number.

Conclusion

Several challenges confront the maximum exploitation of the various tax administrative data sources. For Statistics Canada, the GIFI data set represents a major advance. GIFI for corporations is now becoming a fait accompli, although it has certainly and perhaps inevitably had its share of birth pains. GIFI for unincorporated businesses is scheduled for reference year 2002.

The first statistical GIFI file is scheduled for delivery to PIPES in November 1999. Much effort and energy is being expended preparing for this milestone. The stakes are high. One question remains, "Will the expected GIFI content materialize?". Put in layman terms "the proof will be in the pudding".

Information Sources

- Paper on the Project to Improve Provincial Economic Statistics from the Joint IASS/IAOS Conference. Statistics for Economic and Social Development, September, 1998 D. Royce.
- Unified Enterprise Statistics Program. Project to Improve Provincial Economic Statistics. Technical Series Number 1
- Project to Improve Provincial Economic Statistics. Technical Series Number 5. D. Royce, F. Hardy, G. Beelan.
- Realizing and Measuring Quality Improvements in Provincial Economic Statistics. Paper for the Joint Statistics Meetings. Dallas, Texas – August 1998. P. Smith.

For further information on the PIPES:

Ms. Bonnie Bercik

Co-ordination and Communications

Project to Improve Provincial

Economic Statistics (PIPES)

Statistics Canada

120 Parkdale Avenue

13th floor, Section B7

Ottawa, Ontario K1A 0T6 CANADA

Tel: (613) 951-6790

Fax: (613) 951-0411

e-mail: Bonnie.Bercik@statcan.ca