

Methodology Statement Template – Unified Enterprise Survey Program, Reference Year 2000

Real Estate Agents, Brokers, Appraisers and Other Real Estate Activities – Concepts, Methodology and Data Quality

The following information is to ensure a clear understanding of the basic concepts that are being measured, the underlying survey methodology (how the concepts are measured), and key aspects of data quality. This information will provide you with a better understanding of the strengths and limitations of the data, and of how they can be effectively used and analysed. The information may be of particular importance when making comparisons with data from other surveys or sources of information, and in drawing conclusions regarding changes over time.

Definitions

A **business entity** is an economic transactor having the responsibility and the authority to allocate resources in the production of goods and services.

A **statistical establishment** is one production entity or the smallest grouping of production entities which produces as homogenous a set of goods and/or services as possible, which does not cross provincial boundaries, and for which records provide data on the value of output together with the cost of principal intermediate inputs used and cost and quantity of labour resources used to produce the output.

A **statistical enterprise** is an organisational unit of a business that directs and controls the allocation of resources relating to its domestic operations, and for which consolidated financial and balance sheet accounts are maintained from which international transactions, an international investment position and a consolidated financial position for the unit can be derived.

A **complex enterprise** is an enterprise that operates in more than one province and/or operates in more than one industry and /or is multi-legal (links to more than one legal entity).

A **simple enterprise** is an enterprise that operates in one province, one industry and is a single legal entity.

In complex businesses, there may be an organisational unit above the establishment but below the enterprise. The **statistical company** is defined as the organisational unit for which income and expenditure accounts and balance sheets are maintained from which operating profit and the rate of return on capital can be derived.

Survey Design

The target population consists of all statistical establishments (sometimes referred to as firms or units) classified as Real Estate Agents, Brokers, Appraisers and Other Real Estate Activities according to the North American Industry Classification System (NAICS) during the reference year 2000. Data users who wish to learn more about NAICS, its underlying principles, and many of the other statistical concepts discussed in this brief summary, are referred to the Introduction section of the Statistics Canada publication “North American Industry Classification System: Canada 1997” (catalogue number 12-501).

Even though the basic objective of the survey is to produce estimates for the whole industry—all incorporated and unincorporated businesses—not all businesses are surveyed. The data come from three different sources: a sample of all incorporated businesses with revenue above or equal to a certain threshold (Note: the threshold varies between surveys and sometimes between provinces in the same survey); administrative data for the incorporated businesses with revenue below the specified threshold and finally administrative data to estimate for unincorporated businesses. It should be noted that only basic information is obtained from administrative sources; i.e., total revenue, expenses, depreciation and salaries, wages and benefits. Detailed characteristics such as client base, revenue by type of service, and detailed expense items are collected only for surveyed establishments.

The **frame** is the list of establishments from which the portion eligible for sampling is determined and the sample is taken. The frame provides basic information about each firm, including: address, industry classification, and

information from administrative data sources (as discussed above). The frame is maintained by Statistics Canada's Business Register, and is updated using administrative data.

Prior to the selection of a random sample, establishments are classified into homogeneous groups (i.e., groups with the same NAICS codes, same geography (province/territory), and same business type (incorporated/unincorporated) attributes). Quality requirements are targeted, and then each group is divided into sub-groups called strata: take-all, must-take, and take-some.

The take-all stratum includes the largest firms in terms of performance (based on revenue) in an industry. Every firm is sampled, which means each firm represents itself and is given a weight of one. The must-take stratum is also comprised of self-representing units, but these are selected on the basis of complex structure characteristics (multi-establishment, multi-legal, multi-NAICS, or multi-province enterprises). Units in the take-some strata are subjected to simple random sampling.

Finally, the sample size is increased, mostly to compensate for firms that no longer belong in the industry; i.e., they have gone out of business, changed their primary business activity, they are inactive, or are duplicates on the frame. After removing such firms, the sample size for this survey was 1,261 collection entities.

Collection

Data are collected through a mail-out/mail-back process, while providing respondents with the option of telephone or other electronic filing methods. The statistical establishment is used as the sampling unit, but selected establishments belonging to the same company, the same industry, and the same province are aggregated to create a collection entity. This reduces respondent burden and simplifies collection. Therefore, companies with production in more than one province are mailed one questionnaire per province and instructed to report for all Canadian operations.

Edit and Imputation

Several checks are performed on the collected data to verify internal consistency and identify extreme values. Where information is missing, imputation is performed using either a "nearest neighbour" procedure (donor imputation), using historical data where available or finally, using administrative data as a proxy for reported data.

Estimation

As part of the production of final numbers (referred to as estimation), data for companies operating in more than one province or territory are allocated to the provincial level. Administrative data are used to estimate for the portion of the industry that was excluded from survey activity (i.e. small firms whose revenues fell below cut-off thresholds). Sampled data are then weighted to produce estimates representative of the target population.

Prior to publication, combined survey results are analyzed for comparability; in general, this includes a detailed review of: individual responses (especially for the largest companies), general economic conditions, historic trends, and comparisons with administrative data (e.g., income tax, goods and services tax, payroll deductions records, industry and trade association sources).

Data Quality

While considerable effort is made to ensure high standards throughout all stages of collection and processing, the resulting estimates are inevitably subject to a certain degree of error. These errors can be broken down into two major types: non-sampling and sampling.

Non-sampling error is not related to sampling and may occur for many reasons. For example, non-response is an important source of non-sampling error. Population coverage, differences in the interpretation of questions, incorrect information from respondents, and mistakes in recording, coding and processing data are other examples of non-sampling errors.

The **response rate** for this survey was 61% in reference year 2000, after taking into account the factors discussed in this statement.

Sampling error occurs because population estimates are derived from a sample of the population rather than the entire population. Sampling error depends on factors such as sample size, sampling design, and the method of estimation. An important property of probability sampling is that sampling error can be computed from the sample itself by using a statistical measure called the coefficient of variation (CV). The assumption is that over repeated surveys, the relative difference between a sample estimate and the estimate that would have been obtained from an enumeration of all units in the universe would be less than twice the CV, 95 times out of 100. The range of acceptable data values yielded by a sample is called a confidence interval. Confidence intervals can be constructed around the estimate using the CV. First, we calculate the standard error by multiplying the sample estimate by the CV. The sample estimate plus or minus twice the standard error is then referred to as a 95% confidence interval.

For the Real Estate Agents, Brokers, Appraisers and Other Real Estate Activities, CVs were calculated for each estimate. Generally, the more commonly reported variables obtained very good CVs (10% or less), while the less commonly reported variables were associated with higher but still acceptable CVs (under 25%). The CVs are available upon request.