

Methodology Report on the Geomatics Industry Survey, 2004

1. Introduction

Statistics Canada (STC), in partnership with Natural Resources Canada (NRCan), conducted a survey for reference year 2004 to produce new statistical information on the firms that produce geomatics products or provide geomatics services in Canada. The data from this survey will be used by Natural Resources Canada to determine the competitive position of the industry and its needs in terms of federal government policies and programs.

2. Target population and collection results

The target population for this survey was all Canadian establishments involved in the production of geomatics products or the delivery of geomatics services. The frame constructed for the survey includes all establishments on the Business Register (BR) coded to either NAICS 541360 (Geophysical Surveying and Mapping Services) or NAICS 541370 (Surveying and Mapping (except Geophysical) Services). Additional units added to frame included establishments indicating geomatics activity as part of the Survey of Innovation 2003. Finally, NRCan provided a list of establishments believed to provide geomatic services or produce geomatic products.. After unduplicating these different sources, a total of 4979 establishments were on the survey frame. A pre-contact was made before mailing the questionnaire to determine if the establishments were active and in-scope. During the pre-contact some establishments were determined to be out-of-scope, out-of-business or duplicates and therefore nothing was mailed to them. A questionnaire was sent to 4202 establishments, i.e. all establishments on the survey frame minus the ones found to be out-of-scope through the pre-contact. From the 4202 establishments, 1598 (38%) establishments were unable to be located or contacted, 1249 (30%) completed the questionnaire, 113 (3%) establishments refused to complete it, but were confirmed to be in-scope establishments, 1074 (26%) establishments were found to be out-of-scope or out-of-business and 168 (4%) establishments were other types of out-of-scope units. We have to note that the out-of-scope units were mainly found in Alberta. In fact, 40% of out-of-scope units found during the pre-contact and 60% of out-of-scope units found with the questionnaire were in Alberta. Overall, the survey response rate was 42%. This is a typical response rate for Statistics Canada business surveys – the targeted response rate for this survey was 40%.

3. Data processing

For all establishments who completed the questionnaire, the data was captured and checked for errors and inconsistencies. For questionnaire revenue information, comparisons were made with administrative tax data for the geomatic sector that was analyzed by the Services Industry Division of Statistics Canada. In some cases, it was necessary to recontact the respondents to clarify these inconsistencies. Missing data were then replaced with consistent values (were imputed) using the procedure DONOR in the BANFF system. This procedure uses a nearest neighbour approach to find, for each establishment which required imputation, the valid establishment that is most similar to it and which imputes all required variables of the establishment by transferring the corresponding values from the nearest neighbour record. Note that for the expenses variable, instead of transferring the corresponding values from the nearest neighbour record to the establishment which required imputation, we used the ratio between the

expenses and the total revenue of the donor and applied this ratio to the establishment which required expenses imputation. The highest imputation rates are for the revenues and expenses variables and for the questions related to the obstacles for R&D and exports. The imputation rate for the revenue variable is 20%, the rate for expenses variable is 42% and the rates for the obstacles to R&D and exports are 25% and 36% respectively. For the rest of the variables, the imputation rates are lower than 10%.

For the 2004 reference year, Service Industries Division produced an industry estimate based on a sample of administrative and historical data for establishments classified coded to NAICS 541360 and NAICS 541370.

The Geomatics Industry Survey, 2004, produced 2 estimates based on a census of establishments having any geomatics activity (all known establishments)

- an industry estimate for establishments coded to NAICS 541360 and 541370, comparable to the administrative results of Service Industries Division
- an activity estimate for establishments having any geomatics activity

A comparison between the industry estimates produced by the two projects yielded comparable revenue results at a national level (administrative data result = \$1.95 billion compared to survey result = \$1.69 billion).

The difference in revenues for these 2 NAICS is due to several factors:

1. The administrative data project had special treatment for complex enterprises. For example, in the case of some large and complex enterprises, enterprise-level data were used when the BR did not properly reflect the firm's operations.
2. During the collection phase of the survey, many establishments were coded out-of-scope (i.e. their NAICS code on the BR was incorrect). While this had the effect of improving the quality of the survey results, the administrative data project included all such establishments as being in-scope.
3. Minor variances between the survey and administrative processes:
 - Survey versus administrative tax data. For example, the survey collected gross revenues, whereas the revenue figure on the tax files was operating revenues.
 - Census survey versus sampling of administrative tax data. Whereas the survey attempted to collect information from all in-scope establishments, the administrative data project was based on a sampling of some in-scope records.
 - Survey reweighting for non-response versus using micro-record imputation for administrative tax data. For the survey, non-responding records were dealt with by adjusting the weights assigned to the responding records, such that one responding record might also represent other non-responding units with similar characteristics (i.e. size, province, industry) as the responding record. However, the administrative data project dealt with non-response by attempting to impute the missing information for the entire record.

While there is more variance at the provincial level, especially in Alberta (industry estimates: administrative data result, = \$1.19 billion compared to survey result = \$902 million approximately), this is not surprising. The lower the level of aggregation, the more likely differences are to be found. SBSS and Services looked at all the big firms to make sure the estimates are reasonable. Still, the differences between the two datasets are generally not significant and those errors that can be found are explainable. Statistics Canada is satisfied that the results of the two projects are consistent.

4. Estimation

Since only 1249 establishments completed the questionnaire, the individual values were weighted to represent the whole industry. The weight equals to the number of establishments in the geomatic population that each establishment who completed the questionnaire represents. The value for each respondent establishment was multiplied by the weight for that establishment, and then the weighted data from all respondent establishments belonging to a given estimation domain (e.g. NAICS 541360 in Ontario) were summed to obtain the estimate.

To calculate the weights, all establishments in the population were first grouped in homogeneous groups called strata. The establishments are grouped (stratified) by province, four types of activity groups (NAICS 541330, 541360, 541370 and others) and two size groups (largest size establishments and smallest ones) if possible. The size was measured by the annual gross business income obtained from the BR (Business Register). Because a minimum of about 10 units per stratum (including one or two out-of-scope units) was desired, it was not possible to make a stratum for each combination of province, industry and size. Some strata were grouped to have the minimum. Something similar was tried for each province, although efforts were made to have a separate stratum for the NAICS 541360 and 541370 for each province. For some provinces when the annual gross business income was not well correlated to the gross revenue obtained from geomatic survey, it was decided to not use the size group to stratify the province or alternately to put the biggest establishments according to the survey in a separate stratum.

The weights were calculated within each stratum. The weights are the product of two terms. The first term corresponds to the non-contact adjustment and the second corresponds to the non-response adjustment. The non-contact adjustment represents an adjustment made to compensate all the establishments that were never possible to contact and therefore it is do not known if they are in-scope establishment or not. The non-response adjustment represents an adjustment made to compensate all establishments that refused to completed the questionnaire but it is known that they are in-scope establishment (by the pre-contact result for example). The non-contact adjustment equals the number of establishments in the stratum divided by the sum of the number of establishments in the stratum who completed the questionnaire, the number of establishments in the stratum who refused to complete the questionnaire (but it is known that they are in-scope) and the number of out-of-scope establishments in the stratum. The non-response adjustment equals the sum of the number of establishments in the stratum who completed the questionnaire and the number of establishments in the stratum who refused to complete the questionnaire (but it is known that they are in-scope) divided the number of establishments in the stratum who completed the questionnaire. All the estimates were calculated using the Generalized Estimation Systems (GES) software. As a measure of error, the coefficients of variation for all estimates were calculated by GES.