

Appendix C.

Methodology

SOURCES OF THE DATA

The construction sector includes approximately 650,000 establishments that were determined to be in-scope of the 2002 Economic Census — Construction. This number includes those industries in the North American Industry Classification System (NAICS) definition of construction with at least one paid employee in 2002.

Establishments in the 2002 Economic Census are divided into those sent report forms and those not sent report forms. The coverage of and the method of obtaining census information from each are described below:

1. Establishments sent a report form:

Sample frame establishments. The sample frame consisted of the entire construction universe; there were no subpopulations that were explicitly removed from the sample frame. The sample frame was compiled from a list of all construction companies in the active records of the Internal Revenue Service (IRS) and the Social Security Administration (SSA) that are subject to the payment of Federal Insurance Contributions Act taxes. Under special arrangements, to safeguard their confidentiality, the U.S. Census Bureau obtains information on the location and classification of the companies, as well as their payroll and receipts data from these sources. Unfortunately, these sources do not provide establishment level information for companies with multiple locations. For multilocation companies, the establishment level information is directly obtained from the U.S. Census Bureau's Company Organization Survey. For single-location companies, the IRS-SSA information is generally sufficient for assigning the company to a specific six-digit NAICS industry code.

The 2002 NAICS structure for the construction sector was significantly revised from the 1997 NAICS structure. Initially, only a small proportion of the establishments in the sample frame could be directly assigned a 2002 NAICS industry code with a high degree of confidence. Therefore, a special classification card was mailed to 150,000 construction establishments in early 2002. The goal of this classification card was to obtain the current NAICS industry code prior to assembly of the sample frame for the economic census — construction sample.

2. Establishments not sent a report form:

a. Nonsample frame establishments. There were a limited number of establishments included in the business register who were completely unclassified at the time of the economic census — construction sampling operation. These establishments were mailed a general classification card in early 2003. A portion of these were ultimately determined to be in-scope of the economic census — construction. Since this determination was not made until after the sample selection operation had been completed; these establishments were treated as a supplement to the original universe and were sampled independently for inclusion in the derived estimates.

b. All nonemployers, i.e., all firms subject to federal income tax, with no paid employees, were also excluded from the 2002 sample frame, as in previous censuses. Nonemployers with significant levels of receipts data were identified and included in the census mailout under the presumption that the nonemployer status may have been incorrect. Those determined to have employees are included in this report. Data for nonemployers are not included in this report, but are released in the annual *Nonemployer Statistics* series.

The report forms used to collect information for establishments in this sector are available at help.econ.census.gov/econhelp/resources/.

A more detailed examination of census methodology is presented in the *History of the Economic Census* at www.census.gov/econ/www/history.html.

INDUSTRY CLASSIFICATION OF ESTABLISHMENTS

The classifications for all establishments covered in the 2002 Economic Census — Construction are classified in 1 of 31 industries in accordance with the industry definitions in the *North American Industry Classification System (NAICS), United States, 2002* manual. Changes between 1997 and 2002 affecting this sector are discussed in the text at the beginning of this report. Tables at www.census.gov/epcd/naics02/n02ton97.htm identify those industries that changed between the 1997 North American Industry Classification System (NAICS) and 2002 NAICS.

In the NAICS system, an industry is generally defined as a group of establishments that use similar processes or have similar business activities. To the extent practical, the system uses supply-based or production-oriented concepts in defining industries. The resulting group of establishments must be significant in terms of number, value added by construction, value of business done, and number of employees.

The coding system works in such a way that the definitions progressively become narrower with successive additions of numerical digits. In the construction sector for 2002, there are 3 subsectors (three-digit NAICS), 10 industry groups (four-digit NAICS), 28 NAICS industries (five-digit NAICS) that are comparable with Canadian and Mexican classification, and 31 U.S. industries (six-digit NAICS).

ESTABLISHMENT BASIS OF REPORTING

The 2002 Economic Census — Construction is conducted on an establishment basis. A construction establishment is defined as a relatively permanent office or other place of business where the usual business activities related to construction are conducted. With some exceptions, a relatively permanent office is one that has been established for the management of more than one project or job and that is expected to be maintained on a continuing basis. Such establishment activities include, but are not limited to, estimating, bidding, purchasing, supervising, and operation of the actual construction work being conducted at one or more construction sites. Separate construction reports were not required for each project or construction site.

Companies with more than one construction establishment were required to submit a separate report for each establishment operated during any part of the census year. The construction sector figures represent a tabulation of records for individual establishments, rather than for companies.

If an establishment was engaged in construction and one or more distinctly different lines of economic activity at the same place of business, it was requested to file a separate report for each activity, provided that the activity was of substantial size and separate records were maintained. If a separate establishment report could not be prepared for each activity, then a construction report was requested covering all activities of that establishment providing that the value of construction work exceeded the gross receipts from each of its other activities.

The 2002 Economic Census — Construction excludes data for central administrative offices (CAOs). These would include separately operated administrative offices, warehouses, garages, and other auxiliary units that service construction establishments of the same company. These data are published in a separate report series.

DESCRIPTION OF THE SAMPLE FRAME

The major objective of the sample design was to provide a sample that would provide reliable estimates at the state by industry level. For sample efficiency considerations, the establishments in the initial 2002 construction frame were partitioned into two components for developing estimates within the sample frame. The details of each are described below:

1. Probability-proportionate-to-size (pps) sample. There were three non-overlapping strata for sample selection. An independent sample was selected within each state by industry cell. The details of each stratum were defined as:

-
- Stratum 1. This stratum was comprised of approximately 12,000 establishments associated with multilocation companies. The establishments of these multiunit companies were included in the construction sample with certainty.
 - Stratum 2. This stratum was comprised of approximately 145,000 single-location companies that could be classified into a valid 2002 NAICS industry. These cases accounted for approximately 75 percent of the payroll associated with single-location companies in construction. The industry code for most of these establishments was determined from the special classification card that was mailed in early 2002. This group was partitioned into state by NAICS (six-digit) cells and an independent sample selected from each cell. Within each cell, a probability-proportionate-to-size (pps) sampling strategy was used. Under this approach, the probability of selection for the sample for larger establishments is higher than for smaller establishments. There were approximately 80,000 establishments selected from this group.
 - Stratum 3. This stratum was comprised of the remaining single-location companies. For these companies, we did not have an updated 2002 NAICS industry code. The most recent classification information available for these companies was their 1997 NAICS. Using this 1997 NAICS industry code, this stratum was partitioned into state by NAICS (four-digit) cells; and an independent sample selected from each cell. Again, probability-proportionate-to-size sampling methodology was utilized. There were approximately 30,000 establishments selected from this group.

Subsequent to the initial census mail-out, companies that initiated operations in 2002 were identified via administrative sources. To assure proper representation of the entire in-scope population, simple random samples of these new operations were selected and mailed separately.

2. Estimation and variances. Based on the response data, establishments were assigned to the appropriate NAICS (six-digit) industry. At each level of tabulation, unbiased estimates were derived by summing the weighted establishment data where the establishment sample weight was equal to the inverse of its probability of selection for the construction sample.

The resulting estimates were generated from one of many possible samples and are subject to sampling variability. Estimates of this sample variability were independently derived at all levels of aggregation. These sampling variances were then aggregated to the publication levels for the computation of the relative standard errors.

RELIABILITY OF DATA

The estimates developed from the sample can differ somewhat from the results of a survey covering all companies in the sample lists, but are otherwise conducted under essentially the same conditions as the actual sample survey. The estimates of the magnitude of the sampling errors (the difference between the estimates obtained and the results theoretically obtained from a comparable, complete-coverage survey) are provided by the standard errors of estimates.

The particular sample selected for the construction sector is one of many similar probability samples that, by chance, might have been selected under the same specifications. Each of the possible samples would yield somewhat different sets of results, and the standard errors are measures of the variation of all the possible sample estimates around the theoretically, comparable, complete-coverage values.

Estimates of the standard errors have been computed from the sample data. They are presented in the form of relative standard errors that are the standard errors divided by the estimated values to which they refer.

In conjunction with its associated estimate, the relative standard error may be used to define confidence intervals, or ranges that would include the comparable, complete-coverage value for specified percentages of all the possible samples.

The complete-coverage value would be included in the range:

-
- From one standard error below to one standard error above the derived estimate for about two-thirds of all possible samples.
 - From two standard errors below to two standard errors above the derived estimate for about 19 out of 20 of all possible samples.
 - From three standard errors below to three standard errors above the derived estimate for nearly all samples.

An inference is that the comparable complete-survey result would fall within the indicated ranges and the relative frequencies shown. Those proportions, therefore, may be interpreted as defining the confidence that the estimates from a particular sample would differ from complete-coverage results by as much as one, two, or three standard errors, respectively.

For example, suppose an estimated total is shown at 50,000 with an associated relative standard error of 2 percent, that is, a standard error of 1,000 (2 percent of 50,000). There is approximately 67 percent confidence that the interval 49,000 to 51,000 includes the complete-coverage total, about 95 percent confidence that the interval 48,000 to 52,000 includes the complete-coverage total, and almost certain confidence that the interval 47,000 to 53,000 includes the complete-coverage total.

In addition to the sample errors, the estimates are subject to various response and operational errors: errors of collection; reporting; coding; transcription; imputation for nonresponse, etc. These operational errors also would occur if a complete canvass were to be conducted under the same conditions as the survey. Explicit measures of their effects generally are not available. However, it is believed that most of the important operational errors were detected and corrected during the U.S. Census Bureau's review of the data for reasonableness and consistency. The small operational errors usually remain. To some extent, they are compensating in the aggregated totals shown. When important operational errors were detected too late to correct the estimates, the data were suppressed or were specifically qualified in the tables.

As derived, the estimated standard errors included part of the effect of the operational errors. The total errors, which depend upon the joint effect of the sampling and operational errors, are usually of the order of size indicated by the standard error, or moderately higher. However, for particular estimates, the total error may considerably exceed the standard errors shown. Any figures shown in the tables of this publication having an associated standard error exceeding 75 percent may be combined with higher level totals, creating a broader aggregate, which then may be of acceptable reliability.

DUPLICATION IN VALUE OF CONSTRUCTION WORK

The aggregate of value of construction work reported by all construction establishments in each of the industry, geographic area, or other groupings contains varying amounts of duplication. This is because the construction work of one firm may be subcontracted to other construction firms and may also be included in the subcontractors' value of construction work. Also, part of the value of construction results from the use of products of nonconstruction industries as input materials. These products are counted in the nonconstruction industry, as well as part of the value of construction. Value added avoids this duplication and is, for most purposes, the best measure for comparing the relative economic importance of industries or geographic areas. Value added for construction industries is defined as the dollar value of business done less costs for construction work subcontracted to others and payments for materials, components, supplies, and fuels.

DISCLOSURE

In accordance with federal law governing census reports (Title 13 of the United States Code), no data are published that would disclose the operations of an individual establishment or company. However, the number of establishments in a specific industry or geographic area is not considered a disclosure; therefore, this information may be released even though other information is withheld. Techniques employed to limit disclosure are discussed at www.census.gov/epcd/ec02/disclosure.htm.