

Appendix C.

Methodology

SOURCES OF THE DATA

The manufacturing sector includes approximately 350,000 establishments. This number includes those industries in the North American Industry Classification System (NAICS) definition of manufacturing. The amount of information requested from manufacturing establishments was dependent upon a number of factors. The more important considerations were the size of the company and whether it was included in the Annual Survey of Manufactures (ASM).

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:

- a. ASM sample establishments. This group accounts for approximately 15 percent of all manufacturing establishments. The ASM panel covers all the units of large manufacturing establishments, as well as a sample of the medium and smaller establishments. The probability of selection was proportionate to size. For more information, see the Description of the ASM Survey Sample.

In an economic census year, the ASM report form (MA-10000) replaces the first page of the regular census form for those establishments included in the ASM. In addition to information on employment, payroll, and other items normally requested on the regular census form, establishments in the ASM sample were requested to supply additional information on gross book value of assets and capital expenditures. ASM establishments were also requested to provide information on retirements, depreciation, rental payments, and supplemental labor costs. For establishments not included in the ASM, these additional items were estimated using relationships observed in the ASM establishment data. The census statistics for these variables are a sum of the ASM establishment data and the estimated data for non-ASM establishments. ASM establishments were also requested to provide information for selected purchased services. The census statistics for the purchased service items were derived solely from the ASM establishments. See Appendix A. Explanation of Terms, for an explanation of these items. The census part of the report form is 1 of 220 versions containing product, material, and special inquiries. The diversity of manufacturing activities necessitated the use of this many forms to canvass the 473 manufacturing industries. Each form was developed for a group of related industries.

Appearing on each form was a list of products primary to the group of related industries, as well as secondary products and miscellaneous services that establishments classified in these industries were likely to perform. Respondents were requested to identify the products, the value of each product, and, in certain cases, the quantity of the product shipped during the survey year. Space also was provided for the respondent to describe products not specifically identified on the form.

The report form also contained a materials-consumed inquiry, which varied from form to form depending on the industries being canvassed. The respondents were asked to review a list of materials generally used in their production processes. From this list, each establishment was requested to identify those materials consumed during the survey year, the cost of each, and, in certain cases, the quantity consumed. Once again, space was provided

for the respondent to describe significant materials not listed on the form.

A wide variety of special inquiries were included to measure activities peculiar to a given industry, such as operations performed and equipment used.

- b. Large and medium establishments (non-ASM). Approximately 30 percent of all manufacturing establishments were included in this group. A variable cutoff, based on administrative-record payroll data and determined on an industry-by-industry basis, was used to select those establishments that were to receive 1 of the 220 economic census — manufacturing regular forms. The first page, requesting establishment data for items such as employment and payroll, was standard but did not contain the detailed statistics included on the ASM form. The product, material, and special inquiry sections supplied were based on the historical industry classification of the establishment.
- c. Small single-establishment companies (non-ASM). This group includes approximately 15 percent of all manufacturing establishments. For those industries where application of the variable cutoff for administrative-record cases resulted in a large number of small establishments being included in the mail canvass, an abbreviated “short form” was used. These establishments received 1 of the 31 versions of the short form, which requested summary product and material data and totals but no details on employment, payroll, cost of materials, inventories, and capital expenditures.

Use of the short form has no adverse effect on published totals for the industry statistics, because the same data were collected on the short form as on the long form. However, detailed information on products and materials consumed was not collected on the short form; thus, its use would increase the value of the “not specified by kind” (nsk) categories.

2. Establishments not sent a report form:

- a. Small single-establishment companies not sent a report form. Approximately 40 percent of the manufacturing establishments were small single-establishment companies that were excused from filing a census report. Selection of these establishments was based on two factors: annual payroll and the Census Bureau’s ability to assign the correct six-digit NAICS industry classification to the establishment. For each six-digit NAICS industry code, an annual payroll cutoff was determined. These cutoffs were derived so that the establishments with payroll less than the cutoff were expected to account for no more than 3 percent of the value of shipments for the industry. Generally, all single-establishment companies with less than 5 employees were excused, while all establishments with more than 20 employees were mailed forms. Establishments below the cutoff that could not be directly assigned a six-digit NAICS code were mailed a classification report that requested information for assigning NAICS industry codes. Establishments below the cutoff that could be directly assigned a six-digit NAICS code were excused from filing any report. For below cutoff establishments, information on the physical location, payroll, and receipts was obtained from the administrative records of other federal agencies under special arrangements that safeguarded their confidentiality.

Estimates of data for these small establishments were developed using industry averages in conjunction with the administrative information. The value of shipments and cost of materials were not distributed among specific products and materials for these establishments, but were included in the product and material “not specified by kind” (nsk) categories.

The industry classification codes included in the administrative-record files were assigned on the basis of brief descriptions of the general activity of the establishment. As a result, an indeterminate number of establishments were erroneously coded to a six-digit NAICS industry. This was especially true whenever there was a relatively fine line of demarcation between industries or between manufacturing and nonmanufacturing activity.

Sometimes the administrative-record cases had only two- or three-digit NAICS group classification codes available in the files. For manufacturing, these establishments were sent a

separate classification form, which requested information on the products and services of the establishment. This form was used to code many of these establishments to the appropriate six-digit NAICS level. Establishments that did not return the classification form were coded later to those six-digit NAICS industries identified as “All other” industries within the given subsector.

As a result of these situations, a number of small establishments may have been misclassified by industry. However, such possible misclassification has no significant effect on the statistics, other than on the number of companies and establishments.

The total establishment count for individual industries should be viewed as an approximation rather than a precise measurement. The counts for establishments with 20 employees or more are far more reliable than the count of total number of establishments.

- b. All nonemployers, i.e., all firms subject to federal income tax, with no paid employees, during 2002 are excluded as in previous censuses. 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 — Manufacturing are classified in 1 of 473 industries in accordance with the industry definitions in the *North American Industry Classification System (NAICS), United States, 2002* manual. There were no changes between the 2002 edition and the 1997 edition affecting this sector. When applicable, Appendix F of this report shows the product class and product comparability between the two systems for data in this report.

In the NAICS system, an industry is generally defined as a group of establishments that have similar production processes. 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 manufacture, value of shipments, 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 manufacturing sector for 2002, there are 21 subsectors (three-digit NAICS), 86 industry groups (four-digit NAICS), 184 NAICS industries (five-digit NAICS) that are comparable with Canadian and Mexican classification, and 473 U.S. industries (six-digit NAICS). Product classes and products of the manufacturing industries have been assigned codes based on the industry from which they originate. There are 1,450 product classes (seven-digit codes), 5,674 census products, and an additional 3,746 ten-digit product codes. The ten-digit products are considered the primary products of the industry with the same first six digits.

For the 2002 Economic Census — Manufacturing, all establishments were classified in particular industries based on the products they produced. If an establishment made products of more than one industry, it was classified in the industry with the largest product value. For 2002, there were no “resistance rules” or “frozen industries.”

In ASM years, establishments included in the ASM sample with certainty weights are reclassified by industry only if the change in the primary activity from the prior year is significant or if the change has occurred for 2 successive years. This procedure prevents reclassification when there are minor shifts in product mix.

In ASM years, establishments included in the ASM sample with noncertainty weight are not shifted from one industry classification to another. They are retained in the industry where they were classified in the base census year. However, in the following census year, these ASM plants are allowed to shift from one industry to another.

The results of these rules covering the switching of plants from one industry classification to another are that some industries comprise different mixes of establishments in different survey years. Hence, comparisons between prior-year and current-year published totals, particularly at the six-digit NAICS level, should be viewed with caution. This is particularly true for the comparison between the data shown for a census year versus the data shown for the previous ASM year.

As previously noted, the small establishments that may have been misclassified by industry are usually administrative-record cases whose industry codes were assigned on the basis of incomplete descriptions of the general activity of the establishment. Such possible misclassifications have no significant effect on the statistics other than on the number of companies and establishments.

Establishments frequently make products classified both in their industry (primary products) and other industries (secondary products). Industry statistics (employment, payroll, value added by manufacture, value of shipments, etc.) reflect the activities of the establishments that may make both primary and secondary products. Product statistics, however, represent the output of all establishments without regard for the classification of the producing establishment. For this reason, when relating the industry statistics, especially the value of shipments, to the product statistics, the composition of the industry's output should be considered.

The extent to which industry and product statistics may be matched with each other is measured by the primary product specialization ratio and the coverage ratio. The primary product specialization ratio is the proportion of industry shipments accounted for by the primary products of establishments classified in the industry. The coverage ratio is the proportion of product shipments accounted for by establishments classified in the industry.

ESTABLISHMENT BASIS OF REPORTING

The 2002 Economic Census — Manufacturing is conducted on an establishment basis. A company operating at more than one location is required to file a separate report for each location or establishment. The ASM also is conducted on an establishment basis, but separate reports are filed for just those establishments selected in the sample. Companies engaged in distinctly different lines of activity at one location are requested to submit separate reports, if the plant records permit such a separation and if the activities are substantial in size.

In 2002, as in earlier years, a minimum size limit was set for inclusion of establishments in the census. All establishments employing one person or more at any time during the census year are included. The same size limitation has applied since 1947 in censuses and annual surveys of manufactures. In the 1939 and earlier censuses, establishments with less than \$5,000 value of products were excluded. The change in the minimum size limit in 1947 does not appreciably affect the historical comparability of the census figures, except for data on number of establishments for a few industries.

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

DESCRIPTION OF THE ASM SURVEY SAMPLE

The ASM sample is drawn for the second survey year after a census. The most recent sample was drawn for the 1999 survey year based on the 1997 Economic Census — Manufacturing. This sample will be in place through the 2003 ASM.

In 1997, there were approximately 370,000 individual manufacturing establishments. For sample efficiency and cost considerations, the establishments in the 1997 manufacturing population were partitioned into two components for developing estimates within the ASM. The details of each are described below:

1. Mail stratum. The mail stratum of the survey is comprised of larger single-location manufacturing companies and all manufacturing establishments of multiunit companies (companies

that operate at more than one physical location). Approximately 200,000 of the 370,000 establishments in the 1997 census were assigned to the mail stratum. On an annual basis, the mail stratum is supplemented with larger, newly active single-location companies identified from a list provided by the Internal Revenue Service (IRS) and new manufacturing locations of multiunit companies identified from the Census Bureau's Company Organization Survey (COS).

For the 1999 survey, a new sample of approximately 58,000 individual establishments was selected from the mail stratum assembled from the 1997 census. Supplemental samples representing both 1998 and 1999 births (newly active establishments that were not included in the 1997 census) were also selected. Establishments selected for the sample are mailed an ASM survey questionnaire for each year through 2003.

The 1999-2003 ASM sample design is similar to the one used since 1984. Companies in the 1997 Economic Census — Manufacturing with manufacturing shipments of at least \$500 million were defined as company certainties. For these large companies, each manufacturing establishment is included in the mail sample. For the 1999-2003 sample, there are approximately 500 certainty companies collectively accounting for over 18,000 establishments.

For the remaining portion of the mail component of the survey, the establishment was defined as the sample unit. All establishments with 250 employees or more were defined as employment certainties. Across these arbitrary certainty classes, there were approximately 25,000 establishments included in the sample with certainty. Collectively, these certainty establishments accounted for approximately 80 percent of the total value of shipments in the 1997 Economic Census — Manufacturing.

Smaller establishments in the remaining portion of the mail stratum were sampled with probabilities ranging from .02 to 1.00. The initial probabilities of selection assigned to these establishments were proportionate to a measure-of-size determined for each establishment. The measure-of-size was a function of the establishment's 1997 industry classification and its 1997 product class data. For each product class (1,755) and six-digit industry (473), a desired reliability constraint was specified. Using a technique developed by Dr. James R. Chromy of the Research Triangle Institute, the initial establishment probabilities were optimized such that the expected sample satisfied all industry and product class reliability constraints, while the sample size was minimized. This technique reduces the likelihood of selecting nonrepresentative samples for individual product classes or industries.

This method of assigning probabilities based on product class shipments is motivated by the Census Bureau's primary desire to produce reliable estimates of both product class and industry shipments. The high correlation between shipments and employment, value-added, and other general statistics assures that these variables will also be well represented by the sample. The actual sample selection procedure uses an independent chance of selection method (Poisson sampling) that permits us to prevent small establishments from being selected in consecutive samples without introducing a bias into the survey estimates.

2. Nonmail stratum. The initial nonmail component of the survey was comprised of approximately 170,000 small, single-establishment companies that were tabulated as administrative records in the 1997 Economic Census — Manufacturing. The nonmail stratum is also supplemented annually using the list of newly active single-location companies provided by the Internal Revenue Service (IRS) and payroll cutoffs. Companies with payroll below the payroll cutoff are added to the nonmail stratum. For this portion of the population, sampling is not used. The data for this group are estimated based on selected information obtained annually from the administrative records of the IRS and Social Security Administration (SSA). This administrative information, which includes payroll, total employment, industry classification, and physical location, is obtained under conditions which safeguard the confidentiality of both tax and census records.

RELIABILITY OF DATA

All data compiled in the economic census are subject to nonsampling errors. Nonsampling errors can be attributed to many sources during the development or execution of the census. The following are two ways that further explain this method: ASM Estimating Procedure. Most of the ASM

estimates derived for the mail stratum are computed using a difference estimator. At the establishment level, there is a strong correlation between the current-year data values and the corresponding 1997 (base) data values. Therefore, within the mailed stratum, for each item at each level of aggregation, an estimate of the “difference” between the current year and the base year is computed from sample cases and added to the corresponding base-year values. For the 1998-2002 ASM estimates, the 1997 Economic Census — Manufacturing values serve as the base year. For the 2003 ASM, the base will be updated to be the 2002 Economic Census — Manufacturing.

Due to the positive year-to-year correlation, estimates derived using this methodology are generally more reliable than comparable estimates developed from the current sample data alone. Estimates for the capital expenditures variables are not generated using the difference estimator because the year-to-year correlations are considerably weaker. The standard linear estimator is used for these variables.

For the nonmail stratum, estimates for payroll and employment are directly tabulated from the administrative-record data provided by IRS and SSA. Estimates of data other than payroll and employment are developed from industry averages. Although the nonmail stratum contained approximately 170,000 individual establishments in 1999, it accounts for less than 2 percent of the estimate for total value of shipments at the total manufacturing level.

Corresponding estimates for the mail and nonmail components are combined to produce the estimates included in this publication. ASM Data Qualifications. The estimates developed from the sample are apt to differ somewhat from the results of a survey covering all companies in the sample lists, but 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 ASM 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 for selected ASM statistics in this report. They are represented in the form of relative standard errors (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 (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 that the comparable, complete-survey result would be within the indicated ranges would be correct in approximately 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 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 in this publication having an associated standard error exceeding 15 percent may be combined with higher level totals, creating a broader aggregate, which then may be of acceptable reliability.

DUPLICATION IN COST OF MATERIALS AND VALUE OF SHIPMENTS

Data for cost of materials and value of shipments include varying amounts of duplication, especially at higher levels of aggregation. This is because the products of one establishment may be the materials of another. The value added statistics avoid this duplication and are, for most purposes, the best measure for comparing the relative economic importance of industries and geographic areas.

VALUE OF INDUSTRY SHIPMENTS COMPARED WITH VALUE OF PRODUCT SHIPMENTS

The 2002 Economic Census — Manufacturing shows value of shipments data for industries and products. In the industry statistics tables and files, these data represent the total value of shipments of all establishments classified in a particular industry. The data include the shipments of the products classified in the industry (primary to the industry), products classified in other industries (secondary to the industry), and miscellaneous receipts (repair work, sale of scrap, research and development, installation receipts, and resales). Value of product shipments shown in the products statistics tables and files represent the total value of all products shipped that are classified as primary to an industry regardless of the classification of the producing establishment.

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.

The disclosure analysis for the industry statistics files is based on the total value of shipments. When the total value of shipments cannot be shown without disclosing information for individual companies, the complete line is suppressed except for capital expenditures. Nonetheless, the suppressed data are included in higher-level totals. A separate disclosure analysis is performed for capital expenditures, which can be suppressed even though value of shipments data are published.