
Appendix III

Limitations of the Data

Introduction—The data presented in this *Statistical Abstract* came from many sources. The sources include not only Federal statistical bureaus and other organizations that collect and issue statistics as their principal activity, but also governmental administrative and regulatory agencies, private research bodies, trade associations, insurance companies, health associations, and private organizations such as the National Education Association and philanthropic foundations. Consequently, the data vary considerably as to reference periods, definitions of terms and, for ongoing series, the number and frequency of time periods for which data are available.

The statistics presented were obtained and tabulated by various means. Some statistics are based on complete enumerations or censuses while others are based on samples. Some information is extracted from records kept for administrative or regulatory purposes (school enrollment, hospital records, securities registration, financial accounts, social security records, income tax returns, etc.), while other information is obtained explicitly for statistical purposes through interviews or by mail. The estimation procedures used vary from highly sophisticated scientific techniques, to crude “informed guesses.”

Each set of data relates to a group of individuals or units of interest referred to as the *target universe* or *target population*, or simply as the *universe* or *population*. Prior to data collection the target universe should be clearly defined. For example, if data are to be collected for the universe of households in the United States, it is necessary to define a “household.” The target universe may not be completely tractable. Cost and other considerations may restrict data collection to a *survey universe* based on some available list, such list may be it of date. This list is called a *survey frame* or *sampling frame*.

The data in many tables are based on data obtained for all population units, a *census*, or on data obtained for only a portion, or *sample*, of the population units. When the data presented are based on a sample, the sample is usually a scientifically selected *probability sample*. This is a sample selected from a list or sampling frame in such a way that every possible sample has a known chance of selection and usually each unit selected can be assigned a number, greater than zero and less than or equal to one, representing its likelihood or probability of selection.

For large-scale sample surveys, the probability sample of units is often selected as a multistage sample. The first stage of a multistage sample is the selection of a probability sample of large groups of population members, referred to as primary sampling units (PSUs). For example, in a national multistage household sample, PSUs are often counties or groups of counties. The second stage of a multistage sample is the selection, within each PSU selected at the first stage, of smaller groups of population units, referred to as secondary sampling units. In subsequent stages of selection, smaller and smaller nested groups are chosen until the ultimate sample of population units is obtained. To qualify a multistage sample as a probability sample, all stages of sampling must be carried out using probability sampling methods.

Prior to selection at each stage of a multistage (or a single stage) sample, a list of the sampling units or sampling frame for that stage must be obtained. For example, for the first stage of selection of a national household sample, a list of the counties and county groups that form the PSUs must be obtained. For the final stage of selection, lists of households, and sometimes persons within the households, have to be compiled in the field. For surveys of economic entities and for

the economic censuses the Bureau generally uses a frame constructed from the Bureau's Business Register. The Business Register contains all establishments with payroll in the United States including small single establishment firms as well as large multi-establishment firms.

Wherever the quantities in a table refer to an entire universe, but are constructed from data collected in a sample survey, the table quantities are referred to as *sample estimates*. In constructing a sample estimate, an attempt is made to come as close as is feasible to the corresponding universe quantity that would be obtained from a complete census of the universe. Estimates based on a sample will, however, generally differ from the hypothetical census figures. Two classifications of errors are associated with estimates based on sample surveys: (1) *sampling error*—the error arising from the use of a sample, rather than a census, to estimate population quantities and (2) *non-sampling error*—those errors arising from nonsampling sources. As discussed below, the magnitude of the sampling error for an estimate can usually be estimated from the sample data. However, the magnitude of the nonsampling error for an estimate can rarely be estimated. Consequently, actual error in an estimate exceeds the error that can be estimated.

The particular sample used in a survey is only one of a large number of possible samples of the same size which could have been selected using the same sampling procedure. Estimates derived from the different samples would, in general, differ from each other. The *standard error* (SE) is a measure of the variation among the estimates derived from all possible samples. The standard error is the most commonly used measure of the sampling error of an estimate. Valid estimates of the standard errors of survey estimates can usually be calculated from the data collected in a probability sample. For convenience, the standard error is sometimes expressed as a percent of the estimate and is called the relative standard error or *coefficient of variation* (CV). For example, an estimate of 200 units with an estimated standard error of 10 units has an estimated CV of 5 percent.

A sample estimate and an estimate of its standard error or CV can be used to construct interval estimates that have a prescribed confidence that the interval includes the average of the estimates derived from all possible samples with a known probability. To illustrate, if all possible samples were selected under essentially the same general conditions, and using the same sample design, and if an estimate and its estimated standard error were calculated from each sample, then: 1) Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average estimate derived from all possible samples; 2) approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average estimate derived from all possible samples; and 3) approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average estimate derived from all possible samples.

Thus, for a particular sample, one can say with the appropriate level of confidence (e.g., 90 percent or 95 percent) that the average of all possible samples is included in the constructed interval. Example of a confidence interval: An estimate is 200 units with a standard error of 10 units. An approximately 90 percent confidence interval (plus or minus 1.6 standard errors) is from 184 to 216.

All surveys and censuses are subject to nonsampling errors. Nonsampling errors are of two kinds—*random* and *nonrandom*. Random nonsampling errors arise because of the varying interpretation of questions (by respondents or interviewers) and varying actions of coders, keyers, and other processors. Some randomness is also introduced when respondents must estimate. Nonrandom nonsampling errors result from total nonresponse (no usable data obtained for a sampled unit), partial or item nonresponse (only a portion of a response may be usable), inability or unwillingness on the part of respondents to provide correct information, difficulty interpreting questions, mistakes

in recording or keying data, errors of collection or processing, and coverage problems (overcoverage and undercoverage of the target universe). Random nonresponse errors usually, but not always, result in an understatement of sampling errors and thus an overstatement of the precision of survey estimates. Estimating the magnitude of nonsampling errors would require special experiments or access to independent data and, consequently, the magnitudes are seldom available.

Nearly all types of nonsampling errors that affect surveys also occur in complete censuses. Since surveys can be conducted on a smaller scale than censuses, nonsampling errors can presumably be controlled more tightly. Relatively more funds and effort can perhaps be expended toward eliciting responses, detecting and correcting response error, and reducing processing errors. As a result, survey results can sometimes be more accurate than census results.

To compensate for suspected nonrandom errors, adjustments of the sample estimates are often made. For example, adjustments are frequently made for nonresponse, both total and partial. Adjustments made for either type of nonresponse are often referred to as *imputations*. Imputation for total nonresponse is usually made by substituting for the questionnaire responses of the nonrespondents the “average” questionnaire responses of the respondents. These imputations usually are made separately within various groups of sample members, formed by attempting to place respondents and nonrespondents together that have “similar” design or ancillary characteristics. Imputation for item nonresponse is usually made by substituting for a missing item the response to that item of a respondent having characteristics that are “similar” to those of the nonrespondent.

For an estimate calculated from a sample survey, the *total error* in the estimate is composed of the sampling error, which can usually be estimated from the sample, and the nonsampling error, which usually cannot be estimated from the

sample. The total error present in a population quantity obtained from a complete census is composed of only nonsampling errors. Ideally, estimates of the total error associated with data given in the *Statistical Abstract* tables should be given. However, due to the unavailability of estimates of nonsampling errors, only estimates of the levels of sampling errors, in terms of estimated standard errors or coefficients of variation, are available. To obtain estimates of the estimated standard errors from the sample of interest, obtain a copy of the referenced report which appears at the end of each table.

Principal data bases—Beginning below are brief descriptions of 41 of the sample surveys and censuses that provide a substantial portion of the data contained in this *Abstract*.

U.S. DEPARTMENT OF AGRICULTURE, National Agriculture Statistics Service

Census of Agriculture

Universe, Frequency, and Types of Data:

Complete count of U.S. farms and ranches conducted once every 5 years with data at the national, state, and county level. Data published on farm numbers and related items/characteristics.

Type of Data Collection Operation: Complete census for— number of farms; land in farms; estimated market value of land and buildings, agriculture products sold; total cropland; irrigated land; farm operator characteristics; livestock and poultry inventory and sales; and selected crops harvested. Total farm production expenses, machinery and equipment, fertilizer and chemicals, and farm labor are estimated from a sample of farms.

Data Collection and Imputation Procedures: Data collection is by mailing questionnaires to all farmers and ranchers. Nonrespondents are contacted by telephone and correspondence followups. Imputations were made for all nonresponse item/characteristics.

Estimates of Sampling Error: Variability in the estimates is due to the sample selection and estimation for items collected

by sample and census nonresponse estimation procedures. The CVs for national and state estimates are generally very small. Approximately 85 percent response rate.

Other (nonsampling) Errors: Nonsampling errors are due to incompleteness of the census mailing list, duplications on the list, respondent reporting errors, errors in editing reported data, and in imputation for missing data. Evaluation studies are conducted to measure certain non-sampling errors such as list coverage and classification error. Results from the evaluation program for the 1997 census indicate the net under coverage amounted to about 13 percent of the nations total farms.

Sources of Additional Material: U.S. Department of Agriculture (NASS), *1997 Census of Agriculture*, Volume 2, Subject Series—Part 1, *Agriculture Atlas of the U.S.*; Part 2, *Coverage Evaluation*; Part 3, *Rankings of States and Counties*; Part 4, *History*; Part 5, *ZIP Code Tabulation of Selected Items*; and Volume 3 *Special Studies*, Part 1, *Farm and Ranch Irrigation Survey*; Part 2, *Census of Horticultural Specialities*; Part 3, *Census of Aquaculture*.

Basic Area Frame Sample

Universe, Frequency, and Types of Data: June agricultural survey collects data on planted acreage and livestock inventories. The survey also serves to measure list incompleteness and is subsampled for multiple frame surveys.

Type of Data Collection Operation: Stratified probability sample of about 11,000 land area units of about 1 sq. mile (range from 0.1 sq. mile in cities to several sq. miles in open grazing areas). Sample includes 42,000 parcels of agricultural land. About 20 percent of the sample replaced annually.

Data Collection and Imputation Procedures: Data collection is by personal enumeration. Imputation is based on enumerator observation or data reported by respondents having similar agricultural characteristics.

Estimates of Sampling Error: Estimated CVs range from 1 percent to 2 percent for regional estimates to 3 percent to 6 percent for state estimates of major crop acres and livestock inventories.

Other (nonsampling) Errors: Minimized through rigid quality controls on the collection process and careful review of all reported data.

Sources of Additional Material: U.S. Department of Agriculture, SRS, *Scope and Methods of the Statistical Reporting Service*, (name changed to National Agricultural Statistics Service), Miscellaneous Publication No. 1308, September 1983 (revised).

Multiple Frame Surveys

Universe, Frequency, and Types of Data: Surveys of U.S. farm operators to obtain data on major livestock inventories, selected crop acreage and production, grain stocks, and farm labor characteristics; farm economic data and chemical use data.

Type of Data Collection Operation: Primary frame is obtained from general or special purpose lists, supplemented by a probability sample of land areas used to estimate for list incompleteness.

Data Collection and Imputation Procedures: Mail, telephone, or personal interviews used for initial data collection. Mail nonrespondent followup by phone and personal interviews. Imputation based on average of respondents.

Estimates of Sampling Error: Estimated CV for number of hired farm workers is about 3 percent. Estimated CVs range from 1 percent to 2 percent for regional estimates to 3 percent to 6 percent for state estimates of livestock inventories and crop acreage.

Other (nonsampling) Errors: In addition to above, replicated sampling procedures used to monitor effects of changes in survey procedures.

Sources of Additional Material: U.S. Department of Agriculture, SRS, *Scope and Methods of the Statistical Reporting Service*, (name changed to National Agricultural Statistics Service), Miscellaneous Publication No. 1308, September 1983 (revised).

Objective Yield Surveys

Universe, Frequency, and Types of Data: Surveys for data on corn, cotton, potatoes, soybeans, and wheat, to forecast and estimate yields.

Type of Data Collection Operation: Random location of plots in probability sample. Corn, cotton, soybeans, spring wheat, and durum wheat selected in June from Basic Area Frame Sample (see above). Winter wheat and potatoes selected from March and June multiple frame surveys, respectively.

Data Collection and Imputation Procedures: Enumerators count and measure plant characteristics in sample fields. Production measured from plots at harvest. Harvest loss measured from post harvest gleanings.

Estimates of Sampling Error: CVs for national estimates of production are about 2-3 percent.

Other (nonsampling) Errors: In addition to above, replicated sampling procedures used to monitor effects of changes in survey procedures.

Sources of Additional Material: U.S. Department of Agriculture, SRS, *Scope and Methods of the Statistical Reporting Service*, (name changed to National Agricultural Statistics Service), Miscellaneous Publication No. 1308, September 1983 (revised).

U.S. CENSUS BUREAU

County Business Patterns

Universe, Frequency, and Types of Data: Annual tabulation of basic data items extracted from the Business Register, a file of all known single and multi establishment companies maintained and updated by the Census Bureau. Data include number of establishments, number of employees, first quarter and annual payrolls, and number of establishments by employment size class. Data are excluded for self-employed persons, domestic service workers, railroad employees, agricultural production workers, and most government employees.

Type of Data Collection Operation: The annual Company Organization Survey provides individual establishment data for multi establishment companies. Data for single establishment companies are obtained from various Census Bureau programs, such as the Annual Survey of

Manufactures and Current Business Surveys, as well as from administrative records of the Internal Revenue Service and the Social Security Administration.

Estimates of Sampling Error: Not Applicable. Other (nonsampling) Error: Response rates of greater than 85 percent for the 2000 Company Organization Survey.

Sources of Additional Materials: U.S. Census Bureau, *General Explanation of County Business Patterns*.

Census of Manufactures

Universe, Frequency, and Types of Data: Conducted every 5 years to obtain information on labor, materials, capital input and output characteristics, plant location, and legal form of organization for all plants in the United States with one or more paid employees. Universe was 365,000 manufacturing establishments in 1997.

Type of Data Collection Operation: Complete enumeration of data items obtained from 210,000 firms. Administrative records from Internal Revenue Service (IRS) and Social Security Administration (SSA) are used for 155,000 smaller single-location firms, which were determined by various cutoffs based on size and industry.

Data Collection and Imputation Procedures: Five mail and telephone follow-ups for larger nonrespondents. Data for small single-location firms (generally those with fewer than 10 employees) not mailed census questionnaires were estimated from administrative records of IRS and SSA. Data for nonrespondents were imputed from related responses or administrative records from IRS and SSA. Approximately 8 percent of total value of shipments was represented by fully imputed records in 1987.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: Based on evaluation studies, estimates of non-sampling errors for 1972 were about 1.3 percent for estimated total payroll; 2 percent for total employment; and 1 percent for value of shipments. Estimates for later years are not available.

Sources of Additional Material: U.S. Census Bureau, 1987 *Census of Manufactures, Industry Series, Geographic Area Series, and Subject Series.*

Foreign Trade—Export Statistics

Universe, Frequency, and Types of Data:

The export declarations collected by Customs are processed each month to obtain data on the movement of U.S. merchandise exports to foreign countries. Data obtained include value, quantity, and shipping weight of exports by commodity, country of destination, Customs district of exportation, and mode of transportation.

Type of Data Collection Operation: Shipper's Export Declarations (paper and electronic) are required to be filed for the exportation of merchandise valued over \$1,500 - \$2,500. Customs officials collect and transmit the documents to the Census Bureau on a flow basis for data compilation. Value data for shipments valued under \$2,501 are estimated, based on established percentages of individual country totals.

Data Collection and Imputation Procedures: Statistical copies of Shipper's Export Declarations are received on a daily basis from Customs ports throughout the country and subjected to a monthly processing cycle. They are fully processed to the extent they reflect items valued over \$2,500. Estimates for shipments valued at \$2,500 or less are made, based on established percentages of individual country totals.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: Clerical and complex computer checks intercept most processing errors and minimize otherwise significant reporting errors; other nonsampling errors are caused by undercounting of exports to Canada due to the nonreceipt of some Shipper's Export Declarations.

Sources of Additional Material: U.S. Census Bureau, U.S. *International Trade in Goods and Services: Exports, General Imports, and Imports for Consumption, SITC, Commodity by Country, FT 925* (discontinued after 1996), *U.S. Imports of Merchandise*, and *U.S. Exports of Merchandise.*

Foreign Trade—Import Statistics

Universe, Frequency, and Types of Data:

The import entry documents collected are processed each month to obtain data on the movement of merchandise imported into the United States. Data obtained include value, quantity, and shipping weight by commodity, country of origin, Customs district of entry, and mode of transportation.

Type of Data Collection Operation: Import entry documents, either paper or electronic, are required to be filed for the importation of goods into the United States valued over \$2,000 or for articles which must be reported on formal entries. Customs officials collect and transmit statistical copies of the documents to the Census Bureau on a flow basis for data compilation. Estimates for shipments valued under \$2,001 and not reported on formal entries are based on estimated established percentages of individual country totals.

Data Collection and Imputation Procedures: Statistical copies of import entry documents, received on a daily basis from Customs ports of entry throughout the country, are subjected to a monthly processing cycle. They are fully processed to the extent they reflect items valued at \$2,501 and over or items which must be reported on formal entries.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: Verification of statistical data reporting by Customs officials prior to transmittal and a subsequent program of clerical and computer checks are utilized to hold nonsampling errors arising from reporting and/or processing errors to a minimum.

Sources of Additional Material: U.S. Census Bureau, U.S. *International Trade in Goods and Services: Exports, General Imports, and Imports for Consumption, SITC, Commodity by Country, FT 925* (discontinued after 1996), *U.S. Imports of Merchandise*, and *U.S. Exports of Merchandise.*

Census of Governments

Universe, Frequency, and Types of Data:

Survey of all governmental units in the United States conducted every 5 years

to obtain data on government revenue, expenditures, debt, assets, employment and employee retirement systems, property values, public school systems, and number, size, and structure of governments.

Type of Data Collection Operation: Complete census. List of units derived through classification of government units recently authorized in each state and identification, counting, and classification of existing local governments and public school systems.

Data Collection and Imputation Procedures: Data collected through field and office compilation of financial data from official records and reports for states and large local governments; mail canvass of selected data items, like state tax revenue and employee retirement systems; and collection of local government statistics through central collection arrangements with state governments.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: Some nonsampling errors may arise due to possible inaccuracies in classification, response, and processing.

Sources of Additional Material: U.S. Census Bureau, *Census of Governments, 1997*, various reports, and *State Government Finances: 1992*, GF 92, No. 3.

Annual Surveys of State and Local Government

Universe, Frequency, and Types of Data: Sample survey conducted annually to obtain data on revenue, expenditure, debt, and employment of state and local governments. Universe is all governmental units in the United States (about 87,500).

Type of Data Collection Operation: Sample survey includes all state governments, county governments with 100,000+ population, municipalities with 75,000+ population, townships with 50,000+ population, all school districts with 10,000+ enrollment in March 1999, and other governments meeting certain criteria; probability sample for remaining units.

Data Collection and Imputation Procedures: Field and office compilation of data from official records and reports

for states and large local governments; central collection of local governmental financial data through cooperative agreements with a number of state governments; mail canvass of other units with mail and telephone follow-ups of nonrespondents. Data for nonresponses are imputed from previous year data or obtained from secondary sources, if available.

Estimates of Sampling Error: State and local government totals are generally subject to sampling variability of less than 3 percent.

Other (nonsampling) Errors: Nonresponse rate is less than 15 percent for local governments. Other possible errors may result from undetected inaccuracies in classification, response, and processing.

Sources of Additional Material: U.S. Census Bureau, *Public Employment in 1992*, GE 92, No. 1, *Governmental Finances in 1991-1992*, GF 92, No. 5, and *Census of Governments, 1997*, various reports. Web site references: Employment - state and local site: <<http://www.census.gov/govs/www/apes.html>>. Finance - state and local site: <<http://www.census.gov/govs/estimate.html>>. Finance - state site: <<http://www.census.gov/govs/www/state.html>>.

American Housing Survey

Universe, Frequency, and Types of Data: Conducted nationally in the fall in odd numbered years to obtain data on the approximately 116 million occupied or vacant housing units in the United States (group quarters are excluded). Data include characteristics of occupied housing units, vacant units, new housing and mobile home units, financial characteristics, recent mover households, housing and neighborhood quality indicators, and energy characteristics.

Type of Data Collection Operation: The national sample was a multistage probability sample with about 53,000 units eligible for interview in 2001. Sample units, selected within 394 PSUs, were surveyed over a 4-month period.

Data Collection and Imputation Procedures: For 2001, the survey was conducted by personal interviews. The interviewers obtained the information

from the occupants or, if the unit was vacant, from informed persons such as landlords, rental agents, or knowledgeable neighbors.

Estimates of Sampling Error: For the national sample, illustrations of the S.E. of the estimates are provided in the Appendix D of the 2001 report. As an example, the estimated CV is about 0.2 percent for the estimated percentage of owner occupied units with two persons.

Other (nonsampling) Errors: Response rate was about 93 percent. Nonsampling errors may result from incorrect or incomplete responses, errors in coding and recording, and processing errors. For the 2001 national sample, approximately 1.9 percent of the total housing inventory was not adequately represented by the AHS sample.

Sources of Additional Material: U.S. Census Bureau, *Current Housing Reports*, Series H-150 and H-170, *American Housing Survey*.

Monthly Survey of Construction

Universe, Frequency, and Types of Data: Survey conducted monthly of newly constructed housing units (excluding mobile homes). Data are collected on the start, completion, and sale of housing. (Annual figures are aggregates of monthly estimates.)

Type of Data Collection Operation: Probability sample of housing units obtained from building permits selected from 17,000 places. For nonpermit places, multistage probability sample of new housing units selected in 169 PSUs. In those areas, all roads are canvassed in selected enumeration districts.

Data Collection and Imputation Procedures: Data are obtained by telephone inquiry and field visit.

Estimates of Sampling Error: Estimated CV of 3 percent to 4 percent for estimates of national totals, but may be higher than 20 percent for estimated totals of more detailed characteristics, such as housing units in multiunit structures.

Other (nonsampling) Errors: Response rate is over 90 percent for most items. Nonsampling errors are attributed to definitional problems, differences in

interpretation of questions, incorrect reporting, inability to obtain information about all cases in the sample, and processing errors.

Sources of Additional Material: All data are available on the Internet at <<http://www.census.gov/const/www/newsresconstindex.html>>. Further documentation of the survey is also available at that site.

Value of Construction Put in Place

Universe, Frequency, and Types of Data: Survey conducted monthly on total value of all construction put in place in the current month, both public and private projects. Construction values include costs of materials and labor, contractors' profits, overhead costs, cost of architectural and engineering work, and miscellaneous project costs. (Annual figures are aggregates of monthly estimates.)

Type of Data Collection Operation: Varies by type of activity: Total cost of private one-family houses started each month is distributed into value put in place using fixed patterns of monthly construction progress; using a multistage probability sample, data for private multifamily housing are obtained by mail from owners of multiunit projects. Data for residential additions and alterations are obtained in a quarterly survey measuring expenditures; monthly estimates are interpolated from quarterly data. Estimates of value of private nonresidential construction, and state and local government construction are obtained by mail from owners (or agents) for a probability sample of projects. Estimates of farm nonresidential construction expenditures are based on U.S. Department of Agriculture annual estimates of construction; public utility estimates are obtained from reports submitted to Federal regulatory agencies and from private utility companies; estimates of Federal construction are based on monthly data supplied by Federal agencies.

Data Collection and Imputation Procedures: See "Type of Data Collection Operation." Imputation accounts for approximately 20 percent of estimated value of construction each month.

Estimates of Sampling Error: CV estimates for private nonresidential construction range from 3 percent for estimated value of industrial buildings to 9 percent for religious buildings. CV is approximately 2 percent for total new private nonresidential buildings.

Other (nonsampling) Errors: For directly measured data series based on samples, some nonsampling errors may arise from processing errors, imputations, and misunderstanding of questions. Indirect data series are dependent on the validity of the underlying assumptions and procedures.

Sources of Additional Material: U.S. Census Bureau, *Construction Reports*, Series C30, *Value of Construction Put in Place*.

Annual Survey of Manufactures

Universe, Frequency, and Types of Data:

The Annual Survey of Manufactures (ASM) is conducted annually, except for years ending in 2 and 7 for all manufacturing establishments having one or more paid employees. The purpose of the ASM is to provide key intercensal measures of manufacturing activity, products, and location for the public and private sectors. The ASM provides statistics on employment, payroll, worker hours, payroll supplements, cost of materials, value added by manufacturing, capital expenditures, inventories, and energy consumption. It also provides estimates of value of shipments for 1,500 classes of manufactured products.

Type of Data Collection Operation: The ASM includes approximately 55,000 establishments selected from the census universe of 365,000 manufacturing establishments. Some 25,000 large establishments are selected with certainty, and some 30,000 other establishments are selected with probability proportional to a composite measure of establishment size. The survey is updated from two sources; Internal Revenue Service administrative records are used to include new single-unit manufacturers and the Company organization Survey identifies new establishments of multi-unit forms.

Data Collection and Imputation Procedures: Survey is conducted by mail with phone and mail follow-ups of nonrespondents. Imputation (for all nonresponse items) is based on previous year reports, or for new establishments in survey, on industry averages.

Estimates of Sampling Error: Estimated standard errors for number of employees, new expenditure, and for value added totals are given in annual publications. For U.S. level industry statistics, most estimated standard errors are 2 percent or less, but vary considerably for detailed characteristics.

Other (nonsampling) Errors: Response rate is about 85 percent. Nonsampling errors include those due to collection, reporting, and transcription errors, many of which are corrected through computer and clerical checks.

Sources of Additional Material: U.S. Census Bureau, *Annual Survey of Manufactures*, and Technical Paper 24.

Census of Population

Universe, Frequency, and Types of Data:

Complete count of U.S. population conducted every 10 years since 1790. Data obtained on number and characteristics of people in the U.S.

Type of Data Collection Operation: In 1970, 1980, 1990 and 2000 complete census for some items—age, sex, race, and relationship to householder. In 1970, other items collected from a 5 percent and a 15 percent probability (systematic) sample of the population. In 1980, approximately 19 percent of the housing units were included in the sample; in 1990 and 2000, approximately 17 percent.

Data Collection and Imputation Procedures: In 1970, extensive use of mail questionnaires in urban areas; personal interviews in most rural areas. In 1980, 1990, and 2000, mail questionnaires were used in even more areas than in 1970, with personal interviews in the remainder. Extensive telephone and personal followup for nonrespondents was done in the censuses. Imputations were made for missing characteristics.

Estimates of Sampling Error: Sampling errors for data are estimated for all items collected by sample and vary by

characteristic and geographic area. The CVs for national and state estimates are generally very small.

Other (nonsampling) Errors: Since 1950, evaluation programs have been conducted to provide information on the magnitude of some sources of nonsampling errors such as response bias and undercoverage in each census. Results from the evaluation program for the 1990 census indicate that the estimated net under coverage amounted to about 1.5 percent of the total resident population.

Sources of Additional Material: U.S. Census Bureau, *The Coverage of Population in the 1980 Census, PHC80-E4; Content Reinterview Study: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview, PHC80-E2; 1980 Census of Population, Vol. 1, (PC80-1), Appendixes B, C, and D.*

Current Population Survey (CPS)

Universe, Frequency, and Types of Data: Nationwide monthly sample survey of civilian noninstitutional population, 15 years old or over, to obtain data on employment, unemployment, and a number of other characteristics.

Type of Data Collection Operation: Multi-stage probability sample of about 50,000 households in 754 PSUs in 1996 expanded to about 60,000 households in July 2001. Over-sampling in some states and the largest MSAs to improve reliability for those areas of employment data on annual average basis. A continual sample rotation system is used. Households are in sample 4 months, out for 8 months, and in for 4 more. Month-to-month overlap is 75 percent; year-to-year overlap is 50 percent.

Data Collection and Imputation Procedures: For first and fifth months that a household is in sample, personal interviews; other months, approximately, 85 percent of the data collected by phone. Imputation is done for both item and total nonresponse. Adjustment for total nonresponse is done by a predefined cluster of units, by MSA size and residence; for item nonresponse imputation varies by subject matter.

Estimates of Sampling Error: Estimated CVs on national annual averages for labor force, total employment, and nonagricultural employment, 0.2 percent; for total unemployment and agricultural employment, 1.0 percent to 2.5 percent. The estimated CVs for family income and poverty rate for all persons in 1986 are 0.5 percent and 1.5 percent, respectively. CVs for subnational areas, such as states, would be larger and would vary by area.

Other (nonsampling) Errors: Estimates of response bias on unemployment are not available, but estimates of unemployment are usually 5 percent to 9 percent lower than estimates from reinterviews. Six to 7.0 percent of sample households unavailable for interviews.

Sources of Additional Material: U.S. Census Bureau and Bureau of Labor Statistics, *Current Population Survey; Design and Methodology*, (Tech. Paper 63), available on Internet <www.bls.census.gov/cps/tp/tp63.htm> and Bureau of Labor Statistics, *Employment and Earnings*, monthly, Explanatory Notes and Estimates of Error, Tables 1-A through 1-D and *BLS Handbook of Methods*, Chapter 1 (Bulletin 2490.)

Surveys of Minority- and Women-Owned Business Enterprises (SMOBE/SWOBE)

Universe, Frequency, and Types of Data: The surveys provide basic economic data on businesses owned by Blacks, Hispanics, Asians, Pacific Islanders, Alaska Natives, American Indians, and Women. All firms operating during 1997, except those classified as agricultural, are represented. The lists of all firms (or sample frames) are compiled from a combination of business tax returns and data collected on other economic census reports. The published data include the number of firms, gross receipts, number of paid employees, and annual payroll. The data are presented by geographic area, industry, size of firm, and legal form of organization of firm.

Type of Data Collection Operation: The surveys are based on a stratified probability sample of approximately 2.5 million firms from a universe of approximately 20.8 million firms. There were

approximately 5.3 million firms with paid employees and 15.5 million firms with no paid employees. The data are based on the entire firm rather than on individual locations of a firm.

Data Collection and Imputation Procedures: Data were collected through a mailout/mailback operation. Compensation for missing data is addressed through reweighting, edit correction, and standard statistical imputation methods.

Estimates of Sampling Error: Variability in the estimates is due to the sample selection and estimation for items collected by SMOBE/SWOBE. CVs are applicable to only published cells in which sample cases are tabulated. The CVs for number of firms and receipts at the national level range from 1 to 4 percent. Other (nonsampling) Error: Nonsampling errors are attributed to many sources: inability to obtain information for all cases in the universe, adjustments to the weights of respondents to compensate for nonrespondents, imputation for missing data, data errors and biases, mistakes in recording or keying data, errors in collection or processing, and coverage problems.

Sources of Additional Materials: U.S. Census Bureau, *Guide to the 1997 Economic Census and Related Statistics*.

1997 Economic Census (Geographic Area Series and Subject Series Reports) (for NAICS sectors 22, 42, 4445, 48-49, and 51-81)

Universe, Frequency, and Types of Data: Conducted every 5 years to obtain data on number of establishments, number of employees, total payroll size, total sales, and other industry specific statistics. In 1997, the universe was all employer and nonemployer establishments primarily engaged in wholesale, retail, utilities, finance & insurance, real estate, transportation & warehousing, and other service industries.

Type of Data Collection Operation: All large employer firms were surveyed (i.e. all employer firms above the payroll size cutoff established to separate large from small employers) plus a 5 percent to 25 percent sample of the small employer firms. Firms with no employees were not required to file a census return.

Data Collection and Imputation Procedures: Mail questionnaires were used with both mail and telephone followups for nonrespondents. Data for nonrespondents and for small employer firms not mailed a questionnaire were obtained from administrative records of the IRS and Social Security Administration or imputed. Nonemployer data were obtained exclusively from IRS 1997 income tax returns.

Estimates of Sampling Error: Not applicable for basic data such as sales, revenue, payroll, etc. Other (nonsampling) Errors: Trade area level unit response rates in 1997 ranged from 85 percent to 99 percent. Item response rates ranged from 60 percent to 90 percent with lower rates for the more detailed questions. Nonsampling errors may occur during the collection, reporting, and keying of data, and industry misclassification.

Sources of Additional Material: U.S. Census Bureau, *1997 Economic Census: Geographic Area Series and Subject Series Reports* (by NAICS sector), Appendix C and <www.census.gov/con97.html>

Current Business Surveys

Universe, Frequency, and Types of Data: Provides monthly estimates of retail sales by kind of business, and end-of-month inventories of retail stores; wholesale sales and end-of-month inventories; and annual receipts of selected service industries.

Type of Data Collection Operation: Probability sample of all firms from a list frame and, additionally, for retail and service an area frame. The list frame is the Bureau's Standard Statistical Establishment List (SSEL) updated quarterly for recent birth Employer Identification (EI) Numbers issued by the Internal Revenue Service and assigned a kind-of-business code by the Social Security Administration. The largest firms are included monthly.

Data Collection and Imputation Procedures: Data are collected by mail questionnaire with telephone followups for nonrespondents. Imputation made for each nonresponse item and each item failing edit checks.

Estimates of Sampling Error: For the 2001 monthly surveys, median CVs are about 0.6 percent for estimated total retail sales, 1.4 percent for wholesale sales, 1.7 percent for wholesale inventories. For dollar volume of receipts, CVs from the Service Annual Survey vary by kind of business and range between 1.5 percent to 15.0 percent. Sampling errors are shown in monthly publications.

Other (nonsampling) Errors: Imputation rates are about 18 percent to 23 percent for monthly retail sales, 30 percent for wholesale sales, about 32 percent for monthly wholesale inventories and 14 percent for the *Service Annual Survey*.

Sources of Additional Material: U.S. Census Bureau, *Current Business Reports*, *Monthly Retail Trade*, *Monthly Wholesale Trade*, and *Service Annual Survey*.

Service Annual Surveys

Universe, Frequency, and Types of Data: Provides national estimates for taxable and tax-exempt firms on selected service industries as defined by the North American Industrial Classification System (NAICS). Industries covered by the Service Annual Survey include all or part of the following NAICS sectors: Transportation and Warehouse (NAICS 48-49); Information (NAICS 51); Finance (NAICS 52); Real Estate and Leasing (NAICS 53); Professional, Scientific, and Technical Services (NAICS 54); Administrative and Support and Waste Management and Remediation Services (NAICS 56); Health Care and Social Assistance (NAICS 62); Arts, Entertainment, and Recreation (NAICS 71); and Other Services, except Public Administration (NAICS 81). Items collected include total revenue and revenue from e-commerce transactions; and for selected industries, revenue from detailed service products, total expenses and expenses by major type, revenue from exported services, and inventories. The Service Annual Survey is mailed in January and collects calendar year data for the prior year. Data collection continues for approximately 14 weeks. The data are published approximately 12 months after the initial survey mailing.

Type of Data Collection Operation: Data are obtained from a mail-out/mail-back probability sample of approximately 50,000 selected service firms with paid employees. The survey is supplemented with business births on a quarterly basis. The sampling frame for the Service Annual Survey (SAS) has two types of sampling units represented—Employer Identification Numbers (EINS) and large, multiple-establishment firms. Both sampling units represent clusters of one or more establishments owned or controlled by the same firm. The information used to create these sampling units was extracted from the Census Bureau's Business Register and includes information from the most recent Economic Census.

Estimates of Sampling Error: The estimates are based on a sample. Exact agreement with results that would be obtained from a complete enumeration of firms represented on the sampling frame using the same enumeration procedures is not expected. However, because each firm on the sampling frame has a known probability of being selected into the sample, it is possible to estimate the sampling variability of the survey estimates. The particular sample used in this survey is one of a large number of samples of the same size that could have been selected using the same design. If all possible samples had been surveyed under the same conditions, an estimate of an unknown population value could have been obtained from each sample. These samples give rise to a distribution of estimates for the unknown population value. A statistical measure of the variability among these estimates is the standard error, which can be approximated from any one sample.

Other (nonsampling) Errors: Nonsampling error encompasses all other factors that contribute to the total error of a sample survey estimate and may occur in Census. In the Service Annual Survey, nonsampling error can be attributed to many sources: inability to obtain information about all units in the sample; response errors; differences in the interpretation of the questions; mistakes in coding or keying the data obtained; and other errors of collection, response, coverage, and processing.

Sources of Additional Material: U.S. Census Bureau, *Current Business Reports*, *Service Annual Survey*, Census Bureau Web site <<http://www.census.gov/econ/www/servmenu.html>>.

Wholesale Trade Survey

Universe, Frequency, and Types of Data:

Provides monthly estimates of retail sales by kind of business, and end of month inventories of retail stores; wholesale sales and end of month inventories; and annual receipts of selected service industries.

Type of Data Collection Operation: Probability sample of all firms from a list frame and, additionally, for retail and service an area frame. The list frame is the Bureau's Standard Statistical Establishment List (SSEL) updated quarterly for recent birth Employer Identification (EI) Numbers issued by the Internal Revenue Service and assigned a kind of business code by the Social Security Administration. The largest firms are included monthly.

Data Collection and Imputation Procedures: Data are collected by mail questionnaire with telephone followups for nonrespondents. Imputation made for each nonresponse item and each item failing edit checks.

Estimates of Sampling Error: For the 2001 monthly surveys median CV's are about 0.6 percent for estimated total retail sales, 1.4 for wholesale sales, 1.7 for wholesale inventories. For dollar volume of receipts, CV's from the Service Annual Survey vary by kind of business and range between 1.5 percent to 15.0 percent. Sampling errors are shown in monthly publications.

Other (nonsampling) Errors: Imputation rates are about 18 percent to 23 percent for monthly retail sales, 30 percent for wholesale sales, about 32 percent for monthly wholesale inventories and 14 percent for the Service Annual Survey.

Sources of Additional Material: U.S. Census Bureau, *Current Business Reports*, *Monthly Retail Trade*, *Monthly Wholesale Trade*, and *Service Annual Survey*.

Monthly Retail Trade Survey

Universe, Frequency, and Types of Data:

Provides monthly estimates of retail sales by kind of business and end of month inventories of retail stores; wholesale sales and end of month inventories; and annual receipts of selected service industries.

Type of Data Collection Operation: Probability sample of all firms from a list frame. The list frame is the Bureau's Standard Statistical Establishment List (SSEL) updated quarterly for recent birth Employer Identification (EI) Numbers issued by the Internal Revenue Service and assigned a kind of business code by the Social Security Administration. The largest firms are included monthly; a sample of others is included every month also.

Data Collection and Imputation Procedures: Data are collected by mail questionnaire with telephone followups for nonrespondents. Imputation made for each nonresponse item and each item failing edit checks.

Estimates of Sampling Error: For the 1989 monthly surveys, CV's are about 0.5 percent for estimated total retail sales, 0.99 percent for estimated total retail inventories, 1.2 percent for wholesale sales, 1.7 percent for wholesale inventories. For dollar volume of receipts, CV's from the Service Annual Survey vary by kind of business and range between 1.5 percent to 15.0 percent. Sampling errors are shown in monthly publications.

Other (nonsampling) Errors: Imputation rates are about 20 percent for monthly retail sales, 28 percent for monthly retail inventories, 30 percent for wholesale sales, about 32 percent for monthly wholesale inventories and 14 percent for the Service Annual Survey.

Sources of Additional Material: U.S. Census Bureau, *Current Business Reports*, *Monthly Retail Trade*, *Monthly Wholesale Trade*, *Service Annual Survey*.

**Higher Education General
Information Survey (HEGIS),
Fall Enrollment in Institutions
of Higher Education;** beginning
1986, **Integrated Postsecondary
Education Data Survey (IPEDS),
Fall Enrollment**

Universe, Frequency, and Types of Data: Annual survey of all institutions and branches listed in the *Directory* to obtain data on total enrollment by sex, level of enrollment, type of program, racial/ethnic characteristics (every other year prior to 1989, then annually) and attendance status of student, and on first-time students.

Type of Data Collection Operation: Complete census.

Data Collection and Imputation Procedures: Survey package is usually mailed in the spring with surveys due at varying dates in the summer and fall; mail and phone followup procedures for nonrespondents. Missing data are imputed by using data of similar institutions.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: For degree-granting institutions approximately 96.9 percent response rate in fall 1999.

Sources of Additional Material: U.S. Department of Education, National Center for Education Statistics, *Fall Enrollment in Higher Education*, annual.

**Higher Education General
Information Survey (HEGIS),
Financial Statistics of Institutions
of Higher Education;** beginning 1986, **Integrated Post-
secondary Education Data Sur-
vey (IPEDS), Finance**

Universe, Frequency, and Types of Data: Annual survey of all institutions and branches listed in the *Education Directory, Colleges and Universities* to obtain data on financial status and operations, including current funds revenues, current funds expenditures, and physical plant assets.

Type of Data Collection Operation: Complete census.

Data Collection and Imputation Procedures: Survey package is usually mailed in the spring with surveys due at varying dates in the summer and fall; mail and phone followup procedures for nonrespondents. Missing data are imputed by using data of similar institutions.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: For 1997, 95 percent for degree-granting institutions.

Sources of Additional Material: U.S. Department of Education, National Center for Education Statistics, *Financial Statistics*.

**Higher Education General
Information Survey (HEGIS),
Degrees and Other Formal
Awards Conferred.** Beginning
1986, **Integrated Postsecondary
Education Data Survey (IPEDS),
Completions**

Universe, Frequency, and Types of Data: Annual survey of all institutions and branches listed in the *Education Directory, Colleges and Universities* to obtain data on earned degrees and other formal awards, conferred by field of study, level of degree, sex, and by racial/ethnic characteristics (every other year prior to 1989, then annually).

Type of Data Collection Operation: Complete census.

Data Collection and Imputation Procedures: Survey package is usually mailed in the spring with surveys due at varying dates in the summer and fall; mail and phone followup procedures for nonrespondents. Missing data are imputed by using data of similar institutions.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: For 1999-2000, approximately 84.4 percent response rate for degree-granting institutions.

Sources of Additional Material: U.S. Department of Education, National Center for Education Statistics, *Completions*, annual.

U.S. ENERGY INFORMATION ADMINISTRATION

Residential Energy Consumption Survey

Universe, Frequency, and Types of Data: Quadriennial survey of households and fuel suppliers. Data are obtained on energy-related household characteristics, housing unit characteristics, use of fuels, and energy consumption and expenditures by fuel type.

Type of Data Collection Operation: Probability sample of 5,900 eligible units in 116 PSUs. For responding units, fuel consumption and expenditure data obtained from fuel suppliers to those households.

Data Collection and Imputation Procedures: Personal interviews. Extensive followup of nonrespondents including mail questionnaires for some households. Adjustments for nonrespondents were made in weighting for respondents. Most item nonresponses were imputed.

Estimates of Sampling Error: Estimated CVs for household averages: For consumption, 1.3 percent; for expenditures, 1.0 percent; for various fuels, values ranged from 2.0 percent for electricity to 7.0 percent for LPG.

Other (nonsampling) Errors: Household response rate of 81.0 percent. Nonconsumption data were mostly imputed for mail respondents (2.5 percent of eligible units). Usable responses from fuel suppliers for various fuels ranged from 80.7 percent for electricity to 56.6 percent for fuel oil.

Sources of Additional Material: U.S. Energy Information Administration, *A look at Residential Energy Consumption in 1997*.

U.S. NATIONAL CENTER FOR HEALTH STATISTICS (NCHS)

National Vital Statistics System

Universe, Frequency, and Types of Data: Annual data on births and deaths in the United States.

Type of Data Collection Operation: Mortality data based on complete file of death records, except 1972, based on 50 percent sample. Natality statistics 1951-71,

based on 50 percent sample of birth certificates, except a 20 percent to 50 percent in 1967, received by NCHS.

Beginning 1972, data from some states received through Vital Statistics Cooperative Program (VSCP) and complete file used; data from other states based on 50 percent sample. Beginning 1986, all reporting areas participated in the VSCP.

Data Collection and Imputation Procedures: Reports based on records from registration offices of all states, District of Columbia, New York City, Puerto Rico, Virgin Islands, Guam, American Samoa, and Northern Marianas.

Estimates of Sampling Error: For recent years, CVs for births are small due to large portion of total file in sample (except for very small estimated totals).

Other (nonsampling) Errors: Data on births and deaths believed to be at least 99 percent complete.

Sources of Additional Material: U.S. National Center for Health Statistics, *Vital Statistics of the United States*, Vol. I and Vol. II, annual, and *National Vital Statistics Report*.

National Health Interview Survey (NHIS)

Universe, Frequency, and Types of Data: Continuous data collection covering the civilian noninstitutional population to obtain information on personal and demographic characteristics, illnesses, injuries, impairments, and other health topics.

Type of Data Collection Operation: Multi-stage probability sample of 49,000 households (in 198 PSUs) from 1985 to 1994; 43,000 households (358 design PSUs) from 1995 on, selected in groups of about four adjacent households.

Data Collection and Imputation Procedures: Some missing data items (e.g., race, ethnicity) are imputed using a hot deck imputation value. Unit nonresponse is compensated for by an adjustment to the survey weights.

Estimates of Sampling Error: Estimates of Standard Error (SE): For 1997 medically attended injury or poisoning episodes rates in the past 12 months by falling for: females 47.70 (2.37), and males 36.92 (2.06) per 1,000 population; for

1997 injury episodes rates during the past 12 months inside the home - 29.38 (1.28) per 1,000 population.

Other (nonsampling) Errors: The response rate was 93.8 percent in 1996; in 1999, the total household response rate was 87.6 percent, with the final family response rate of 86.1 percent, and the final sample adult response rate of 69.6 percent; in 2000, the total household response rate was 88.9 percent, with the final family response rate of 72.1 percent, and the final sample adult response rate was 72.1 percent for the NHIS. (Note: The NHIS sample redesign was conducted in 1995, and the NHIS questionnaire was redesigned in 1997.)

Sources of Additional Material: U.S. National Center for Health Statistics, Summary Health Statistics for U.S. Children: National Health Interview Survey, 1997, Vital and Health Statistics, Series 10 #203; U.S. National Center for Health Statistics, Summary Health Statistics for the U.S. Population: National Health Interview Survey, 1997, Vital and Health Statistics, Series 10 #204; U.S. National Center for Health Statistics, Summary Health Statistics for U.S. Adults: National Health Interview Survey, 1997, Vital and Health Statistics, Series 10 #205 (In preparation); U.S. National Center for Health Statistics, Design and Estimation for the National Health Interview Survey, 1995-2004, Vital and Health Statistics, Series 2 #130; U.S. National Center for Health Statistics, Summary Health Statistics Technical Report: National Health Interview Survey, 1997-2003, Vital and Statistics, Series 2 #134.

U.S. BUREAU OF JUSTICE
STATISTICS (BJS)

National Crime Victimization Survey

Universe, Frequency, and Types of Data: Monthly survey of individuals and households in the United States to obtain data on criminal victimization of those units for compilation of annual estimates.

Type of Data Collection Operation: National probability sample survey of about 50,000 interviewed households in 376 PSUs selected from a list of

addresses from the 1980 census, supplemented by new construction permits and an area sample where permits are not required.

Data Collection and Imputation Procedures: Interviews are conducted every 6 months for 3 years for each household in the sample; 8,300 households are interviewed monthly. Personal interviews are used in the first interview; the intervening interviews are conducted by telephone whenever possible.

Estimates of Sampling Error: CVs averaged over the period 1998-2001 are: "3.7 percent for personal crimes (includes all crimes of violence plus purse snatching crimes), 3.8 percent for crimes of violence; 12.1 percent for estimate of rape/sexual assault counts; 7.9 percent for robbery counts; 4.1 percent for assault counts; 11.2 percent for purse snatching (it refers to purse snatching and pocket picking); 2.5 percent for property crimes; 3.8 percent for burglary counts; 2.7 percent for theft (of property); and 5.2 percent for motor vehicle theft counts.

Other (nonsampling) Errors: Respondent recall errors which may include reporting incidents for other than the reference period; interviewer coding and processing errors; and possible mistaken reporting or classifying of events. Adjustment is made for a household noninterview rate of about 7 percent and for a within-household noninterview rate of 10 percent.

Sources of Additional Material: U.S. Bureau of Justice Statistics, *Criminal Victimization in the United States*, annual.

U.S. FEDERAL BUREAU OF
INVESTIGATION

Uniform Crime Reporting (UCR) Program

Universe, Frequency, and Types of Data: Monthly reports on the number of criminal offenses that become known to law enforcement agencies. Data are collected on crimes cleared by arrest, by age, sex, and race of offender, and on assaults on law enforcement officers.

Type of Data Collection Operation: Crime statistics are based on reports of crime data submitted either directly to the FBI

by contributing law enforcement agencies or through cooperating state UCR programs.

Data Collection and Imputation Procedures: States with UCR programs collect data directly from individual law enforcement agencies and forward reports, prepared in accordance with UCR standards, to FBI. Accuracy and consistency edits are performed by FBI.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: Coverage of 94 percent of the population (96 percent in MSA's, 87 percent in "other cities," and 88 percent in rural areas) by UCR program, though varying number of agencies report. Some error may be present through incorrect reporting.

Sources of Additional Material: U.S. Federal Bureau of Investigation, *Crime in the United States*, *Hate Crime Statistics*, annual, *Law Enforcement Officers Killed & Assaulted*, annual.

U.S. BUREAU OF LABOR STATISTICS

U.S. Census Bureau, Current Population Survey (CPS)

Universe, Frequency, and Types of Data: Nationwide monthly sample survey of civilian noninstitutional population, 15 years old or over, to obtain data on employment, unemployment, and a number of other characteristics.

Type of Data Collection Operation: Multi-stage probability sample of about 50,000 households in 754 PSUs in 1996, expanded to about 60,000 households in July 2001. Over-sampling in some states and the largest MSA's to improve reliability for those areas of employment data on annual average basis. A continual sample rotation system is used. Households are in sample 4 months, out for 8 months, and in for 4 more. Month-to-month overlap is 75 percent; year-to-year overlap is 50 percent.

Data Collection and Imputation Procedures: For first and fifth months that a household is in sample, personal interviews; other months, approximately, 85 percent of the data collected by phone. Imputation is done for both item and total nonresponse. Adjustment for total nonresponse is done by a predefined

cluster of units, by MSA size and residence; for item nonresponse imputation varies by subject matter.

Estimates of Sampling Error: Estimated CVs on national annual averages for labor force, total employment, and nonagricultural employment, 0.2 percent; for total unemployment and agricultural employment, 1.0 percent to 2.5 percent. The estimated CVs for family income and poverty rate for all persons in 1986 are 0.5 percent and 1.5 percent, respectively. CVs for subnational areas, such as states, would be larger and would vary by area.

Other (nonsampling) Errors: Estimates of response bias on unemployment are not available, but estimates of unemployment are usually 5 percent to 9 percent lower than estimates from reinterviews. About 7.5 percent of sample households unavailable for interviews.

Sources of Additional Material: U.S. Census Bureau and Bureau of Labor Statistics, *Current Population Survey; Design and Methodology*, (Tech. Paper 63 RV), available on Internet <www.bls.census.gov/cps/tp/tp63.htm> and Bureau of Labor Statistics, *Employment and Earnings*, monthly, Explanatory Notes and Estimates of Error, Tables 1-A through 1-D and *BLS Handbook of Methods*, Chapter 1 (Bulletin 2490).

Consumer Price Index (CPI)

Universe, Frequency, and Types of Data: Monthly survey of price changes of all types of consumer goods and services purchased by urban wage earners and clerical workers prior to 1978, and urban consumers thereafter. Both indexes continue to be published.

Type of Data Collection Operation: Prior to 1978, sample of various consumer items in 87 urban areas; thereafter, in 85 PSUs, except from January 1987 through March 1988, when 91 areas were sampled.

Data Collection and Imputation Procedures: Prices of consumer items are obtained from about 50,000 housing units, and 23,000 other reporters in 87 areas. Prices of food, fuel, and a few other items are obtained monthly; prices of most other commodities and services

are collected every month in the three largest geographic areas and every other month in others.

Estimates of Sampling Error: Estimates of standard errors are available.

Other (nonsampling) Errors: Errors result from inaccurate reporting, difficulties in defining concepts and their operational implementation, and introduction of product quality changes and new products.

Sources of Additional Material: U.S. Bureau of Labor Statistics, Internet site <<http://www.stats.bls.gov/cpihome.htm>> and *BLS Handbook of Methods*, Chapter 17, Bulletin 2490. U.S. Bureau of Labor Statistics Internet sites <<http://www.bls.gov/ppi>>.

Producer Price Index (PPI)

Universe, Frequency, and Types of Data: Monthly survey of producing companies to determine price changes of all commodities produced in the United States for sale in commercial transactions. Data on agriculture, forestry, fishing, manufacturing, mining, gas, electricity, public utilities, and a few services.

Type of Data Collection Operation: Probability sample of approximately 30,000 establishments that result in about 100,000 price quotations per month.

Data Collection and Imputation Procedures: Data are collected by mail and facsimile. If transaction prices are not supplied, list prices are used. Some prices are obtained from trade publications, organized exchanges, and government agencies. To calculate index, price changes are multiplied by their relative weights taken from 1997 shipment values from the Census of Manufactures.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: Not available at present.

Sources of Additional Material: U.S. Bureau of Labor Statistics, *BLS Handbook of Methods*, Chapter 14, Bulletin 2490. U.S. Bureau of Labor Statistics Internet sites <<http://www.bls.gov/ppi>>.

Current Employment Statistics (CES) Program

Universe, Frequency, and Types of Data: Monthly survey covering about 7 million establishments to obtain data on employment, hours, and earnings, by industry.

Type of Data Collection Operation: Sample survey of over 300,000 establishments in June 2002.

Data Collection and Imputation Procedures: Cooperating state agencies mail questionnaires to sample establishments to develop state and local estimates; information is forwarded to BLS where national estimates are prepared.

Estimates of Sampling Error: Not available until survey redesign is completed.

Other (nonsampling) Errors: Estimates of employment adjusted annually to reflect complete universe. Average adjustment is 0.3 percent over the last decade.

Sources of Additional Material: U.S. Bureau of Labor Statistics, *Employment and Earnings*, monthly, Explanatory Notes and Estimates of Error, Tables 2-A through 2-H.

National Compensation Survey

Universe, Frequency, and Types of Data: Nationwide sample survey of establishments of all employment size classes, stratified by geographic area, in private industry and state and local government. Data collected include wages and salaries, employer costs of employer compensation, and employee benefits. Data produced include percent changes in the cost of employment cited in the ECI and costs per hour worked for individual benefits cited in the ECEC. The survey provides data by ownership (Private industry and state and local government), industry sector, major industry divisions, major occupational groups, bargaining status, metropolitan area status, and census region. ECEC also provides data by establishment size class.

Type of Data Collection Operation: Probability proportionate to size sample of establishments. The sample is replaced on a continual basis. Establishments are

in the survey for approximately 5 years, with some establishments replaced each quarter.

Data Collection and Imputation Procedures: For the initial visit, data are primarily collected in a personal visit to the establishment. Quarterly updates are obtained primarily by mail, fax, and telephone. Imputation is done for individual benefits.

Estimates of Sampling Error: Because standard errors vary from quarter to quarter, the ECI uses a 5-year moving average of standard errors to evaluate published series. These standard errors are available at <<http://www.bls.gov/ncs/ect/home.htm>>.

Other (nonsampling) Errors: Nonsampling errors have a number of potential sources. The primary sources are (1) survey nonresponse and (2) data collection and processing errors. Nonsampling errors are not measured. Procedures have been implemented for reducing nonsampling errors, however, primarily through quality assurance programs. These programs include the use of data collection reinterviews, observed interviews, computer edits of the data, and systematic professional review of the reports on which the data are recorded. The programs also serve as a training device to provide feedback to the field economists, or data collectors, on errors. And, they provide information on the sources of error which can be remedied by improved collection instructions or computer processing edits. Extensive training of field economists is also conducted to maintain high standards in data collection.

Sources of Additional Material: Bureau of Labor Statistics, *BLS Handbook of Methods*, Chapter 8 (Bulletin 2490) and <<http://www.bls.gov/ncs>>.

BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM

Survey of Consumer Finances

Universe, Frequency, and Types of Data: Periodic sample survey of families. In this survey a given household is divided into a primary economic unit and other

economic units. The primary economic unit, which may be a single individual, is generally chosen as the unit that contains the person who either holds the title to the home or is the first person listed on the lease. The primary unit is used as the reference family. The survey collects detailed data on the composition of family balance sheets, the terms of loans, and relationships with financial institutions. It also gathered information on the employment history and pension rights of the survey respondent and the spouse or partner of the respondent.

Type of Data Collection Operation: The survey employs a two-part strategy for sampling families. Some families were selected by standard multistage area-probability sampling methods from the 48 contiguous states. The remaining families in the survey were selected using tax data under the strict rules governing confidentiality and the rights of potential respondents to refuse participation.

Data Collection and Imputation Procedures: The Survey Research Center at the University of Michigan collected the 1989 survey data between August 1989 and March 1990. Adjustments for nonresponse errors are made through systematic imputation of unanswered questions and through weighting adjustments based on data used in the sample design for families that refused participation.

Estimates of Sampling Error: Because of the complex design of the survey, the estimation of potential sampling errors is not straightforward.

Other (nonsampling) Errors: The achieved sample of 3,143 families represents a response rate of about 69 percent in the area-probability sample and a rate of about 34 percent in the tax-data sample. Proper training of interviewers and careful design of questionnaires were used to control inaccurate survey responses.

Sources of Additional Material: Board of Governors of the Federal Reserve System, "Changes in Family Finances from 1983 to 1989: Evidence from the Survey of Consumer Finances," *Federal Reserve Bulletin*, January 1992.

U.S. INTERNAL REVENUE SERVICE

Statistics of Income, Individual Income Tax Returns

Universe, Frequency, and Types of Data:

Annual study of unaudited individual income tax returns, forms 1040, 1040A, and 1040EZ, filed by U.S. citizens and residents. Data provided on various financial characteristics by size of adjusted gross income, marital status, and by taxable and nontaxable returns. Data by state, based on 100 percent file, also include returns from 1040NR, filed by nonresident aliens plus certain self-employment tax returns.

Type of Data Collection Operation: Annual 1999 stratified probability sample of approximately 177,000 returns broken into sample strata based on the larger of total income or total loss amounts as well as the size of business plus farm receipts. Sampling rates for sample strata varied from 0.05 percent to 100 percent.

Data Collection and Imputation Procedures: Computer selection of sample of tax return records. Data adjusted during editing for incorrect, missing, or inconsistent entries to ensure consistency with other entries on return.

Estimates of Sampling Error: Estimated CVs for tax year 1999: Adjusted gross income less deficit 0.11 percent; salaries and wages 0.21 percent; and tax-exempt interest received 1.78 percent. (State data not subject to sampling error.)

Other (nonsampling) Errors: Processing errors and errors arising from the use of tolerance checks for the data.

Sources of Additional Material: U.S. Internal Revenue Service, *Statistics of Income, Individual Income Tax Returns*, annual.

Statistics of Income, Sole Proprietorship Returns and Statistics of Income Bulletin

Universe, Frequency, and Types of Data:

Annual study of unaudited income tax returns of nonfarm sole proprietorships, form 1040 with business schedules. Data provided on various financial characteristics by industry.

Type of Data Collection Operation: Stratified probability sample of approximately 51,000 sole proprietorships for tax year 1999. The sample is classified based on presence or absence of certain business schedules; the larger of total income or loss; and size of business plus farm receipts. Sampling rates vary from 0.05 percent to 100 percent.

Data Collection and Imputation Procedures: Computer selection of sample of tax return records. Data adjusted during editing for incorrect, missing, or inconsistent entries to ensure consistency with other entries on return.

Estimates of Sampling Error: Estimated CVs for tax year 1999 are available. For sole proprietorships, business receipts, 0.71 percent; net income, (less loss), 1.05 percent; depreciation 1.42 percent.

Other (nonsampling) Errors: Processing errors and errors arising from the use of tolerance checks for the data.

Sources of Additional Material: U.S. Internal Revenue Service, *Statistics of Income, Sole Proprietorship Returns* (for years through 1980) and *Statistics of Income Bulletin*, Vol. 21, No. 1 (summer 2001).

Statistics of Income, Partnership Returns and Statistics of Income Bulletin

Universe, Frequency, and Types of Data:

Annual study of unaudited income tax returns of partnerships, Form 1065. Data provided on various financial characteristics by industry.

Type of Data Collection Operation: Stratified probability sample of approximately 43,000 partnership returns from a population of 2.0 million filed during calendar year 1999. The sample is classified based on combinations of gross receipts, net income or loss, and total assets, and on industry. Sampling rates vary from 0.08 percent to 100 percent.

Data Collection and Imputation Procedures: Computer selection of sample of tax return records. Data are adjusted during editing for incorrect, missing, or inconsistent entries to ensure consistency with other entries on return. Data not available due to regulations are not imputed.

Estimates of Sampling Error: Estimated CVs for tax year 1999 (latest available): For number of partnerships, 0.3 percent; business receipts, 0.2 percent; net income, 0.5 percent; net loss, 1.6 percent.

Other (nonsampling) Errors: Processing errors and errors arising from the use of tolerance checks for the data.

Sources of Additional Material: U.S. Internal Revenue Service, *Statistics of Income, Partnership Returns and Statistics of Income Bulletin*, Vol. 21, No. 2 (fall 2001).

Corporation Income Tax Returns

Universe, Frequency, and Types of Data: Annual study of unaudited corporation income tax returns, Forms 1120 and 1120 (A, F, L, PC, REIT, RIC, and S), filed by corporations or businesses legally defined as corporations. Data provided on various financial characteristics by industry and size of total assets, and business receipts.

Type of Data Collection Operation: Stratified probability sample of approximately 141,400 returns for 1999, distributed by sample classes generally based on type of return, size of total assets, size of net income or deficit, and selected business activity. Sampling rates for sample strata varied from .25 percent to 100 percent.

Data Collection and Imputation Procedures: Computer selection of sample of tax return records. Data adjusted during editing for incorrect, missing, or inconsistent entries to ensure consistency with other entries on return and to achieve statistical definitions.

Estimates of Sampling Error: Estimated CVs for 1999: Number of returns in subgroups ranged from 0.2 percent with assets under \$100,000, to 0 percent with assets over \$100 mil.; for amount of net income 0.22 percent.

Other (nonsampling) Errors: Processing errors and errors arising from the use of tolerance checks for the data.

Sources of Additional Material: U.S. Internal Revenue Service, *Statistics of Income, Corporation Income Tax Returns*, annual.

U.S. SOCIAL SECURITY ADMINISTRATION

Benefit Data

Universe, Frequency, and Types of Data: All persons receiving monthly benefits under Title II of Social Security Act. Data on number and amount of benefits paid by type and state.

Type of Data Collection Operation: Data based on administrative records. Data based on 100 percent files, as well as 10 percent and 1 percent sample files.

Data Collection and Imputation Procedures: Records used consist of actions pursuant to applications dated by subsequent post-entitlement actions.

Estimates of Sampling Error: Varies by size of estimate and sample file size.

Other (nonsampling) Errors: Processing errors, which are believed to be small.

Sources of Additional Material: U.S. Social Security Administration, *Annual Statistical Supplement to the Social Security Bulletin*.

Supplemental Security Income (SSI) Program

Universe, Frequency, and Types of Data: All eligible aged, blind, or disabled persons receiving SSI benefit payments under SSI program. Data include number of persons receiving federally administered SSI, amounts paid, and state administered supplementation.

Type of Data Collection Operation: Data based on administrative records.

Data Collection and Imputation Procedures: Data adjusted to reflect returned checks and overpayment refunds. For federally administered payments, actual adjusted amounts are used.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: Processing errors, which are believed to be small.

Sources of Additional Material: U.S. Social Security Administration, *Annual Statistical Supplement to the Social Security Bulletin*.