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 regis-
tration, 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may restrict data collection to a *survey universe* based on some available list, such list may be inaccurate or out of date.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) *nonsampling 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.

Source of Additional Material: The Federal Committee on Statistical Methodology (FCSM) is an interagency committee dedicated to improving the quality of federal statistics <http://fcsm.ssd.census.gov>.

Principal data bases—Beginning below are brief descriptions of 35 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

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, National Agricultural Statistics Service, USDA's National Agricultural Statistics Service: The Fact Finders of Agriculture, March 2007.

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 follow-up 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, National Agricultural Statistics Service), USDA's National Agricultural Statistics Service: The Fact Finders of Agriculture, March 2007.

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 to 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, National Agricultural Statistics Service), USDA's National Agricultural Statistics Service: The Fact Finders of Agriculture, March 2007.

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 42,000 interviewed households in 203 PSUs selected from a list of addresses from the 1990 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; 7,000 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 for 2005 estimates are: 4.1 percent for personal crimes (includes all crimes of violence plus purse snatching crimes),

4.2 percent for crimes of violence;16.2 percent for estimate of rape/sexual assault counts; 9.7 percent for robbery counts; 4.4 percent for assault counts; 15.1 percent for purse snatching/pocket picking; 2.2 percent for property crimes; 4.2 percent for burglary counts; 2.5 percent for theft (of property); and 6.6 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 9 percent and for a within-household noninterview rate of 16 percent.

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

U.S. Bureau of Labor Statistics

Consumer Expenditure Survey (CE)

Universe, Frequency and Types of Data: Consists of two continuous components: a quarterly interview survey and a weekly diary or recordkeeping survey. They are nationwide surveys that collect data on consumer expenditures, income, characteristics, and assets and liabilities. Samples are national probability samples of households that are representative of the civilian noninstitutional population. The surveys have been ongoing since 1980.

Type of Data Collection Operation: The Interview Survey is a panel rotation survey. Each panel is interviewed for five quarters and then dropped from the survey. About 7,500 consumer units are interviewed each quarter. The Diary Survey sample is new each year and consists of about 7,500 consumer units. Data are collected on an ongoing basis in 102 areas of the country.

Data Collection and Imputation Procedures: For the Interview Survey, data are collected by personal interview with each consumer unit interviewed once per quarter for five consecutive quarters. Designed to collect information that respondents can recall for 3 months or longer, such as large or recurring expenditures. For the Diary Survey, respondents record all their expenditures in a self-reporting diary for two consecutive 1-week periods. Designed to pick up items difficult to recall over a long period, such as detailed food expenditures. Missing or invalid attributes, expenditures, or incomes are imputed. Assets and liabilities are not imputed. The U.S. Census Bureau collects the data for the Bureau of Labor Statistics.

Estimates of Sampling Error: Standard error tables are available since 2000.

Other (nonsampling) Errors: Includes incorrect information given by respondents, data processing errors, interviewer errors, and so on. They occur regardless of whether data are collected from a sample or from the entire population.

Sources of Additional Material: Bureau of Labor Statistics, see Internet site <http://www.bls.gov/cex>.

Consumer Price Index (CPI)

Universe, Frequency, and Types of Data: A 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, and since 1998, sample of various consumer items in 87 urban areas; from 1978–1997, 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 each month from about 23,000 retail outlets and from about 4,000 housing units 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.bls.gov/cpi/home.htm> and *BLS Handbook of Methods*, Chapter 17, see Internet site <ttp://www.bls.gov/opub/hom/pdf/homch17.pdf>.

Current Employment Statistics (CES) Program

Universe, Frequency, and Types of Data: Monthly survey drawn from a sampling frame of over 8 million unemployment insurance tax accounts in order to obtain data by industry on employment, hours, and earnings.

Type of Data Collection Operation: In 2006, the CES sample included about 160,000 businesses and government agencies, which represent approximately 400,000 individual worksites.

Data Collection and Imputation Procedures: Each month, the state agencies cooperating with BLS, as well as BLS Data Collection Centers, collect data through various automated collection modes and mail. BLS Washington staff prepares national estimates of employment, hours, and earnings while states use the data to develop state and area estimates.

Estimates of Sampling Errors: The relative standard error for total nonfarm employment is 0.1 percent. From April 2002 to March 2003, the cumulative net birth/death model added 469,000.

Other (nonsampling) Errors: Estimates of employment adjusted annually to reflect complete universe. Average adjustment is 0.2 percent over the last decade, with an absolute range from less than 0.05 percent to 0.6 percent.

Sources of Additional Material: U.S. Bureau of Labor Statistics, Employment and Earnings, monthly, Explanatory Notes and Estimates of Errors, Tables 2-A through 2-F. See <http://www.bls.gov/web/cestntab.htm>.

National Compensation Survey (NCS)

Universe, Frequency, and Types of Data: NCS collects data from establishments of all employment-size classes in private industries as well as state and local governments. The survey stratifies its data by geographic area and industry. NCS collects data on work schedules, wages, salaries, and employer costs for employee benefits. For approximately 80 metropolitan areas and the nation, NCS produces information on workers' earnings and benefits in a variety of occupations at different work levels. NCS is also responsible for two quarterly releases: the Employment Cost Index (ECI), which measures percent changes in the cost of employment, and the Employer Costs for Employee Compensation (ECEC), which measures costs per hour worked for individual benefits. The survey provides data by industry sector, industry division, occupational group, bargaining status, metropolitan area status, census region, and census division. ECEC also provides data by establishment-size class.

Type of Data Collection Operation: Establishments are selected for the survey based on a probability-proportionate-to-employment technique. NCS replaces its sample on a continual basis. Private industry establishments are in the survey for approximately 5 years.

Data Collection and Imputation Procedures: A personal visit to the establishment is the initial source for collecting data. Communication via mail, fax, and telephone provide quarterly updates. Imputation is done for individual benefits.

Estimates of Sampling Error: NCS uses 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. The use of quality assurance programs reduces the potential for nonsampling errors.

These programs include the use of reinterviews, interview observations, and the systematic professional review of reports. The programs also serve as a training device that provides feedback on errors for field economists (or data collectors). Quality assurance programs also provide information on sources of error. This information is used to improve procedures that result in fewer errors. NCS also conducts extensive training of field economists to maintain high standards in data collection.

Sources of Additional Material: Bureau of Labor Statistics, *BLS Handbook of Methods*, Chapter 8 <http://www.bls.gov/opub/hom/pdf/homch8.pdf>.

Producer Price Index (PPI)

Universe, Frequency, and Types of Data: Monthly survey of producing companies to determine price changes of all commodities and services produced in the United States for sale in commercial transactions. Data on agriculture, forestry, fishing, manufacturing, mining, gas, electricity, construction, public utilities, wholesale trade, retail trade, transportation, healthcare, and other 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 the Census Bureau's 1997 shipment values from their Census of Industries.

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 site <http://stats.bls.gov/ppi>.

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 person or couple who either holds the title to the home or is listed on the lease, along with all other people in the household who are financially dependent on that person or couple. The primary economic 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 gathers 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 are selected by standard multistage area probability sampling methods applied to all 50 states. The remaining families in the survey are selected using statistical records derived from tax returns, under the strict rules governing confidentiality and the rights of potential respondents to refuse participation.

Data Collection and Imputation Procedures: National Opinion Research Center (NORC) at the University of Chicago has collected data for the survey since 1992. Since 1995, the survey has used computer-assisted personal interviewing. Adjustments for nonresponse are made through multiple 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. A replicate-based procedure is available.

Other (nonsampling) Errors: The survey aims to complete 4,500 interviews, with about two thirds of that number deriving from the area-probability sample.

The response rate is typically about 70 percent for the area-probability sample and about 35 percent over all strata in the tax-data sample. Proper training and monitoring of interviewers, careful design of questionnaires, and systematic editing of the resulting data were used to control inaccurate survey responses.

Sources of Additional Material: Board of Governors of the Federal Reserve System, "Recent Changes in U.S. Family Finances: Evidence from the 2001 and 2004 Survey of Consumer Finances," *Federal Reserve Bulletin*, 2006, <http://www.federalreserve.gov/Pubs/Bulletin>.

U.S. CENSUS BUREAU

2002 Economic Census (Industry Series, Geographic Area Series and Subject Series Reports) (for NAICS sectors 22, 42, 44-45, 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/receipts/revenue, and other industry-specific statistics. In 2002, the universe was all employer and nonemployer establishments excluding agriculture, forestry, fishing and hunting, and government.

Type of Data Collection Operation: All large employer firms were surveyed (i.e., all employer firms above payroll-size cutoffs 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 sent a census return.

Data Collection and Imputation Procedures: Mail questionnaires were used with both mail and telephone follow-ups for nonrespondents. Businesses also had the option to respond electronically. Data for nonrespondents and for small employer firms not mailed a questionnaire were obtained from administrative records of other federal agencies or imputed. Nonemployer data were obtained exclusively from IRS 2002 income tax returns.

Estimates of Sampling Error: Not applicable for basic data such as sales, revenue, receipts, payroll, etc. Other (nonsampling) errors: establishment response rates by NAICS sector in 2002 ranged from 80 percent to 89 percent. Item response rates generally ranged from 50 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 due to industry misclassification.

Sources of Additional Material: U.S. Census Bureau, *2002 Economic Census: Industry Series, Geographic Area Series* and *Subject Series Reports* (by NAICS sector), Appendix C and <http://www.census.gov/econ/census02/guide/index.html>.

American Community Survey (ACS)

Universe, Frequency, and Types of Data: Nationwide survey to obtain data about demographic, social, economic, and housing characteristics of people, households, and housing units. Covers household population and excludes the population living in institutions, college dormitories, and other group quarters.

Type of Data Collection Operation: First-phase sampling is performed during both Main and Supplemental sampling for approximately 3,000,000 housing units in the U.S. and 36,000 in Puerto Rico (PR). First stage sampling defines the universe for the second stage of sampling through two steps. First, all addresses that were in a first-stage sample within the past four years are excluded from eligibility. This ensures that no address is in sample more than once in any 5-year period. The second step is to select a 20 percent systematic sample of "new" units, i.e. those units that have never appeared on a previous Master Address File (MAF) extract. Each new address is systematically assigned to either the current year or to one of four back-samples. This procedure maintains five equal partitions of the universe.

Data Collection and Imputation Procedures: The American Community Survey is conducted every month on independent samples. Each housing unit in the

independent monthly samples is mailed a prenotice letter announcing the selection of the address to participate, a survey questionnaire package, and a reminder postcard. These sample units receive a second (replacement) questionnaire package if the initial questionnaire has not been returned by a scheduled date. In the mail-out/mail-back sites, sample units for which a questionnaire is not returned in the mail and for which a telephone number is available are defined as the telephone nonresponse follow-up universe. Interviewers attempt to contact and interview these mail nonresponse cases. Sample units from all sites that are still unresponsive two months after the mailing of the survey questionnaires and directly after the completion of the telephone follow-up operation are subsampled at rates between 1 in 2 and 1 in 3. The selected nonresponse units are assigned to Field Representatives (FRs), who visit the units, verify their existence or declare them nonexistent, determine their occupancy status, and conduct interviews. After data collection is completed, any remaining incomplete or inconsistent information was imputed during the final automated edit of the collected data.

Estimates of Sampling Error: The data in the ACS products are estimates of the actual figures that would have been obtained by interviewing the entire population using the same methodology. The estimates from the chosen sample also differ from other samples of housing units and persons within those housing units.

Other (nonsampling) Errors: Nonsampling Error—In addition to sampling error, data users should realize that other types of errors may be introduced during any of the various complex operations used to collect and process survey data. An important goal of the ACS is to minimize the amount of nonsampling error introduced through nonresponse for sample housing units. One way of accomplishing this is by following up on mail nonrespondents.

Sources of Additional Material: U.S. Census Bureau, American Community Survey Web site available on Internet, <http://www.census.gov/acs/www/index.html>, U.S. Census Bureau, American Community Survey Accuracy of the Data documents available on the Internet, <http://www.census.gov/acs/www/UseData/Accuracy/Accuracy1.htm>.

American Housing Survey

Universe, Frequency, and Types of Data: Conducted nationally in odd numbered years to obtain data on the approximately 121 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 57,000 units eligible for interview in 2005. Sample units, selected within 394 PSUs, were surveyed over a 4-month period.

Data Collection and Imputation Procedures: For 2005, 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 Standard Error (SE) of the estimates are provided in Appendix D of the 2003 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 92 percent. Nonsampling errors may result from incorrect or incomplete responses, errors in coding and recording, and processing errors. For the 2005 national sample, approximately 2.2 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* see <http://www.census.gov/hhes/www/ahs.html>.

Annual Survey of Government Employment and Payroll

Universe, Frequency, and Types of Data: The survey measures the number of state, local, and federal civilian government employees and their gross payrolls for the pay period including March 12, 2005. The survey is conducted annually. The survey provides data on full-time and part-time employment, part-time hours worked, full-time equivalent employment, and payroll statistics by governmental function (elementary and secondary education, higher education, police protection, fire protection, financial administration, central staff services, judicial and legal, highways, public welfare, solid waste management, sewerage, parks and recreation, health, hospitals, water supply, electric power, gas supply, transit, natural resources, correction, libraries, air transportation, water transport and terminals, other education, state liquor stores, social insurance administration, and housing and community development).

Type of Data Collection Operations: The survey sample is taken from the 2002 Census of Governments and contains approximately 11,000 local government units. These units were sampled from a sampling frame that contained 83,767 local governments (county, city, township, special district, school districts) in addition to 50 state governments and the District of Columbia. This frame was slightly different from the Annual Finance Survey sampling frame. Thirty-nine of the state governments provided data from central payroll records for all or most of their agencies/institutions. Data for agencies and institutions for the remaining state governments were obtained by mail canvass questionnaires. Local governments were also canvassed using a mail questionnaire. However, elementary and secondary school system data in Florida, North Dakota, and Washington were supplied by special arrangements with the state education agency in each of these states. All respondents receiving the mail questionnaire had the option of responding using the Employment Web site developed for reporting data. Approximately 22.6 percent of the state agency and local government respondents chose to respond on the Web.

Editing and Imputation Procedures: Editing: Editing is a process that ensures survey data are accurate, complete, and consistent. Efforts are made at all phases of collection, processing, and tabulation to minimize errors. Although some edits are built into the Internet data collection instrument and the data entry programs, the majority of the edits are performed after the case has been loaded into the Census Bureau's database. Edits consist primarily of two types: consistency and a ratio of the current year's reported value to the prior year's value. The consistency edits check the logical relationships of data items reported on the form. For example, if a value exists for employees for a function, then a value must exist for payroll also. If part-time employees and payroll are reported then part-time hours must be reported and vice versa. The current year/prior year edits compare data for the number of employees, the function reported for the employees, and the average salary between reporting years. If data falls out of acceptable tolerance levels, the item is flagged for review. Some additional checks are made comparing data from the Annual Finance Survey to data reported on the Annual Survey of Government Employment and Payroll to verify that employees reported on the Annual Survey of Government Employment and Payroll at a particular function have a corresponding expenditure on the Finance Survey. For both types of edits, the edit results are reviewed by analysts and adjusted when needed. When the analyst is unable to resolve or accept the edit failure, contact is made with the respondent to verify or correct the reported data.

Imputation: Not all respondents answer every item on the questionnaire. There are also questionnaires that are not returned despite efforts to gain a response. Imputation is the process of filling in missing or invalid data with reasonable values in order to have a complete data set. For general purpose governments and for schools, the imputations were based on recent historical

data from either a prior year annual survey or the most recent Census of Governments, if it was available. These data were adjusted by a growth rate that was determined by the growth of units that were similar (in size, geography, and type of government) to the nonrespondent. If there was no recent historical data available, the imputations were based on the data from a randomly selected donor that was similar to the nonrespondent. This donor's data was adjusted by dividing each data item by the population (or enrollment) of the donor and multiplying the result by the nonrespondent's population (or enrollment). For special districts, if prior year data are available, the data are brought forward with a national level growth rate applied. Otherwise, the data are imputed to be zero. In cases where good secondary data sources exist, the data from those sources were used.

Estimates of Sampling Error: Estimated relative standard errors for all variables are given in tabulations on the Web site. For U.S. and state-and-local government-level estimates of total full-time equivalents and total payroll, most relative standard errors are generally less than 1 percent, but vary considerably for detailed characteristics.

Other (nonsampling) Errors: Although every effort is made in all phases of collection, processing, and tabulation to minimize errors, the sample data are subject to nonsampling errors such as inability to obtain data for every variable from all units in the sample, inaccuracies in classification, response errors, misinterpretation of questions, mistakes in keying and coding, and coverage errors. These same errors may be evident in census collections and may affect the Census of Governments data used to adjust the sample during the estimation phase and used in the imputation process.

Sources of Additional Material: <http://www.census.gov/govs/www/apes.html> and <http://www.census.gov/govs/www/apesstl05.html>.

Annual Survey of Government Finances

Universe, Frequency, and Types of Data: The United States Census Bureau conducts an Annual Survey of Government Finances, as authorized by law under Title 13, United States Code, Section 182. Alternatively, every five years, in years ending in a '2' or '7,' a Census of Governments, including a Finance portion, is conducted under Title 13, Section 161. The survey coverage includes all state and local governments in the United States. For both the census and annual survey, the finance detail data is equivalent, encompassing the entire range of government finance activities—revenue, expenditure, debt, and assets.

Type of Data Collection Operations: The data collection phase for the annual survey made use of two methods to obtain data: mail canvass and central collection from state sources. In 28 states, all or part of the general purpose finance data for local governments was obtained from cooperative arrangements between the Census Bureau and a state government agency. These usually involved a data collection effort carried out to meet the needs of both agencies—the state agency for purposes of audit, oversight, or information, and the Census Bureau for statistical purposes. Data for the balance of local governments in this annual survey were obtained via mail questionnaires sent directly to county, municipal, township, special district, and school district governments. School district data were collected via cooperative arrangements with state education agencies. Data for state governments were compiled by analysts of the Census Bureau, usually with the cooperation and assistance of state officials. The data were compiled from state government audits, budgets, and other financial reports, either in printed or electronic format. The compilation generally involved recasting the state financial records into the classification categories used for reporting by the Census Bureau.

Data Collection and Imputation Procedures: Survey is conducted by mail with mail follow-ups of nonrespondents.

Imputation for all nonresponse items is based on previous year reports or, for new governments, on data from similar donors.

Estimates of Sampling Error: The local government statistics in this survey are developed from a sample survey. Therefore, the local totals, as well as national or state and local aggregates, are considered estimated amounts subject to sampling error. State government finance data are not subject to sampling. Consequently, state-local aggregates shown here are more reliable (on a relative standard error basis) than the local government estimates they include. Estimates of major United States totals for local governments are subject to a computed sampling variability of less than one-half of 1 percent. State and local government totals are generally subject to sampling variability of less than 3 percent.

Other (nonsampling) Errors: The estimates are also subject to the inaccuracies in classification, response, and processing. Efforts were made at all phases of collection, processing, and tabulation to minimize errors. However, the data are still subject to errors from estimating for missing data, errors from misreported data, errors from miscoding, and difficulties in identifying every unit that should be included in the report. Every effort was made to keep such errors to a minimum through care in examining, editing, and tabulating the data reported by government officials.

Sources of Additional Material: <http://www.census.gov/govs/www/financegen.html> and <http://www.census.gov/govs/www/05censustechdoc.html>.

Annual Survey of Manufactures (ASM)

Universe, Frequency, and Types of Data: The Annual Survey of Manufactures 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 1,800 classes of manufactured products.

Type of Data Collection Operation: The ASM includes approximately 50,000 establishments selected from the census universe of 346,000 manufacturing establishments. Approximately 24,000 large establishments are selected with certainty, and the remaining 26,000 other establishments are selected with probability proportional to a composite measure of establishment size. The survey is updated from two sources: Internal Revenue Service (IRS) administrative records are used to include new single-unit manufacturers and the Company Organization Survey identifies new establishments of multiunit 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 relative standard errors for number of employees, new expenditures, and for value added totals are given in annual publications. For U.S. level industry statistics, most estimated relative standard errors are 2 percent or less, but vary considerably for detailed characteristics.

Other (nonsampling) Errors: The unit 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 the 1990 and 2000 censuses, the 100 percent items included: age, date of birth,

sex, race, Hispanic origin, and relationship to householder. 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 1980, 1990, and 2000, mail questionnaires were used extensively with personal interviews in the remainder. Extensive telephone and personal follow-up 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 coefficients of variation (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 indicated that the estimated net undercoverage amounted to about 1.5 percent of the total resident population. For Census 2000, the evaluation program indicated a net overcount of 0.5 percent of the 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. *Content Reinterview Survey: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview*, 1990, CPH-E-1; Effectiveness of Quality Assurance, CPH-E-2; Programs to Improve Coverage in the 1990 Census, 1990, CPH-E-3. For Census 2000 evaluations, see <http://www.census.gov/pred/www>.

County Business Patterns

Universe, Frequency, and Types of Data: County Business Patterns is an annual tabulation of basic data items extracted from the Business Register, a file of all known single- and multilocation employer companies maintained and updated by the U.S. 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 individuals, private households, railroad employees, agricultural production workers, and most government employees.

Type of Data Collection Operation: The annual Company Organization Survey provides individual establishment data for multilocation 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 IRS, the Social Security Administration, and the Bureau of Labor Statistics.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: The data are subject to nonsampling errors, such as inability to identify all cases in the universe; definition and classification difficulties; differences in interpretation of questions; errors in recording or coding the data obtained; and estimation of employers who reported too late to be included in the tabulations and for records with missing or misreported data.

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

Current Population Survey (CPS)

Universe, Frequency, and Types of Data: Nationwide monthly sample designed primarily to produce national and state estimates of labor force characteristics of the civilian noninstitutionalized population 16 years of age and older.

Type of Data Collection Operation: Multistage probability sample that currently includes 72,000 households from 824 sample areas. Oversampling in some states to improve data reliability for those areas on an 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: The national total estimates of the civilian labor force and of employment have monthly CVs of about .2 percent and annual average CVs of about .125 percent. Unemployment is a much smaller characteristic and consequently has substantially larger CVs than the civilian labor force or employment. The national unemployment rate, the most important CPS statistic, has a monthly CV of about 2 percent and an annual average CV of about 1 percent. The CVs for states vary since more populous states have larger samples (and smaller CVs) than states with smaller populations. Assuming a 6 percent unemployment rate, the smallest states have monthly CVs of about 17 percent and annual average CVs of about 8 percent. The estimated CVs for family income and poverty rate for all persons in 2005 are .4 percent and 1.2 percent, respectively. CVs for subnational areas, such as states, tend to be larger and vary by area.

Other (nonsampling) Errors: Estimates of response bias on unemployment are available. Estimates of unemployment rate from reinterviews range from −2.4 percent to 1.0 percent of the basic CPS unemployment rate (over a 30-month span from January 2004 through June 2006). Eligible CPS households are approximately 82 percent of the assigned households, with a corresponding response rate of 91 percent.

Sources of Additional Material: U.S. Census Bureau and Bureau of Labor Statistics, Current Population Survey: Design and Methodology, (Technical Paper 66), available on the Internet <http://www.census.gov/prod/2006pubs/tp-66.pdf> and the Bureau of Labor Statistics, <http://www.bls.gov/cps/> and the *BLS Handbook of Methods*, Chapter 1, available on the Internet at <http://www.bls.gov/opub/hom/homch1_a.htm>.

Foreign Trade—Export Statistics

Universe, Frequency, and Types of Data: The export declarations collected by U.S. Bureau of Customs and Border Protection 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, district of exportation, and mode of transportation.

Type of Data Collection Operation: Shipper's Export Declarations (paper and electronic) are generally required to be filed for the exportation of merchandise valued over $2,500. U.S. Bureau of Customs and Border Protection officials collect and transmit the documents to the Census Bureau on a flow basis for data compilation. 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 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: The goods data are a complete enumeration of documents collected by the U.S. Bureau of Customs and Border Protection and are not subject to sampling errors; but they are subject to several types of nonsampling errors. Quality assurance procedures are performed at every stage of collection, processing and tabulation;

however the data are still subject to several types of nonsampling errors. The most significant of these include reporting errors, undocumented shipments, timeliness, data capture errors, and errors in the estimation of low-valued transactions. Additional information on errors affecting export data can be found at <http://www.census.gov/foreign-trade/Press-Release/current_press_release/explain.pdf>.

Sources of Additional Material: U.S. Census Bureau, FT 900 U.S. International Trade in Goods and Services, FT 925 (discounted after 1996), U.S. Merchandise Trade, FT 895 U.S. Trade with Puerto Rico and U.S. Possessions, FT 920 U.S. Merchandise trade: selected highlights, and Information Section on Goods and Services at <http://www.census.gov/ft900>.

Foreign Trade—Import Statistics

Universe, Frequency, and Types of Data: The import entry documents collected by U.S. Bureau of Customs and Border Protection 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, 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. U.S. Bureau of Customs and Border Protection 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 for individual country totals.

Data Collection and Imputation Procedures: Statistical copies of import entry documents, received on a daily basis from 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 $2,001 and over or items which must be reported on formal entries.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: The goods data are a complete enumeration of documents collected by the U.S. Bureau of Customs and Border Protection and are not subject to sampling errors; but they are subject to several types of nonsampling errors. Quality assurance procedures are performed at every stage of collection, processing and tabulation; however the data are still subject to several types of nonsampling errors. The most significant of these include reporting errors, undocumented shipments, timeliness, data capture errors, and errors in the estimation of low-valued transactions. Additional information on errors affecting import data can be found at <http://www.census.gov/foreign-trade/Press-Release/current_press_release/explain.pdf>.

Sources of Additional Material: U.S. Census Bureau, FT 900 U.S. International Trade in Goods and Services, FT 925 (discounted after 1996), U.S. Merchandise Trade, FT 895 U.S. Trade with Puerto Rico and U.S. Possessions, FT920 U.S. Merchandise Trade: selected highlights, and Information Section on Goods and Services at <http://www.census.gov/ft900>.

Monthly Retail Trade and Food Service Survey

Universe, Frequency, and Types of Data: Provides monthly estimates of retail and food service sales by kind of business and end of month inventories of retail stores.

Type of Data Collection Operation: Probability sample of all firms from a list frame. The list frame is the Bureau's Business Register updated quarterly for recent birth Employer Identification (EI) Numbers issued by the IRS 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 follow-ups and

fax reminders for nonrespondents. Imputation is made for each nonresponse item and each item failing edit checks.

Estimates of Sampling Error: For the 2006 monthly surveys, CVs are about 0.4 percent for estimated total retail sales and 0.7 percent for estimated total retail inventories. Sampling errors are shown in monthly publications.

Other (nonsampling) Errors: Imputation rates are about 22 percent for monthly retail and food service sales, and 29 percent for monthly retail inventories.

Sources of Additional Material: U.S. Census Bureau, Current Business Reports, Annual Revision of Monthly Retail and Food Services: Sales and Inventories.

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: A multistage probability sample of approximately 900 of the 20,000 permit-issuing jurisdictions in the U.S. was selected. Each month in each of these permit offices, field representatives list and select a sample of permits for which to collect data. To obtain data in areas where building permits are not required, a multistage probability sample of 70 land areas (census tracts or subsections of census tracts) was selected. All roads in these areas are canvassed and data are collected on all new residential construction found. Sampled buildings are followed up until they are completed (and sold, if for sale).

Data Collection and Imputation Procedures: Data are obtained by telephone inquiry and/or field visit. Nonresponse/undercoverage adjustment factors are used to account for late reported data.

Estimates of Sampling Error: Estimated CV of 3 percent to 4 percent for estimates of national totals of units started, 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/starts>, <http://www.census.gov/newhomesales> or <http://www.census.gov/const/www/newsresconstindex.html>. Further documentation of the survey is also available at those sites.

Nonemployer Statistics

Universe, Frequency, and Types of Data: Nonemployer statistics are an annual tabulation of economic data by industry for active businesses without paid employees that are subject to federal income tax. Data showing the number of firms and receipts by industry are available for the U.S., states, counties, and metropolitan areas. Most types of businesses covered by the Census Bureau's economic statistics programs are included in the nonemployer statistics. Tax-exempt and agricultural-production businesses are excluded from nonemployer statistics.

Type of Data Collection Operation: The universe of nonemployer firms is created annually as a byproduct of the Census Bureau's Business Register processing for employer establishments. If a business is active but without paid employees, then it becomes part of the potential nonemployer universe. Industry classification and receipts are available for each potential nonemployer business. These data are obtained primarily from the annual business income tax returns of the IRS.The potential nonemployer universe undergoes a series of complex processing, editing, and analytical review procedures at the Census Bureau to distinguish nonemployers from employers, and to correct and complete data items used in creating the data tables.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: The data are subject to nonsampling errors, such as industry misclassification as well as errors of response, keying, nonreporting, and coverage.

Sources of Additional Material: U. S. Census Bureau, Nonemployer Statistics <http://www.census.gov/epcd/nonemployer/index.html>.

Service Annual Survey

Universe, Frequency, and Types of Data: The U.S. Census Bureau conducts the Service Annual Survey to provide nationwide estimates of revenues and expenses for selected service industries. Estimates are summarized by industry classification based on the 2002 North American Industry Classification System (NAICS). Selected service industries covered by the Service Annual Survey include all or part of the following NAICS sectors: Transportation and Warehousing (NAICS 48–49); Information (NAICS 51); Finance and Insurance (NAICS 52); Real Estate and Rental 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). Data collected include total revenue, total expenses, detailed expenses, revenue from e-commerce transactions; and for selected industries, revenue from detailed service products, revenue from exported services, and inventories. For industries with a significant nonprofit component, separate estimates are developed for taxable firms and firms and organizations exempt from federal income taxes. Questionnaires are mailed in January and request annual data for the prior year. Estimates are published approximately 12 months after the initial survey mailing.

Type of Data Collection Operation: The Service Annual Survey estimates are developed from a probability sample of employer firms and administrative records for nonemployers. Service Annual Survey questionnaires are mailed to a probability sample that is periodically reselected from a universe of firms located in the United States and having paid employees. The sample includes firms of all sizes and covers both taxable firms and firms exempt from federal income taxes. Updates to the sample are made on a quarterly basis to account for new businesses. Firms without paid employees, or nonemployers, are included in the estimates through imputation and/or administrative records data provided by other federal agencies. Links to additional information about confidentiality protection, sampling error, nonsampling error, sample design, definitions, and copies of the questionnaires may be found on the Internet at <http://www.census.gov/econ/www/servmenu.html>.

Estimates of Sampling Error: CVs for the 2005 Service Annual Survey estimates range from 0.4 percent to 1.8 percent for total revenue estimates computed at the NAICS sector (two-digit NAICS code) level. The full 2005 Service Annual Survey results, including coefficients of variations (CVs), can be found at <http://www.census.gov/econ/www/servmenu.html>. Links to additional information regarding sampling error may be found at: <http://www.census.gov/svsd/www/cv.html>.

Other (Nonsampling) Errors: Data are imputed for unit nonresponse, item nonresponse, and for reported data that fails edits. The percent of imputed data for total revenue for the 2005 Service Annual Survey is approximately 9 percent.

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>.

Survey of Business Owners (SBO)

Universe, Frequency, and Types of Data: The Survey of Business Owners (SBO), formerly known as the Surveys of Minority- and Women-Owned Business Enterprises (SMOBE/SWOBE), provides statistics that describe the composition of U.S. businesses by gender, Hispanic

or Latino origin, and race. Data are presented for businesses owned by American Indians and Alaska Natives, Asians, Blacks, Hispanics, Native Hawaiians and Other Pacific Islanders, and Women. All U.S. firms operating during 2002 with receipts of $1,000 or more, which are classified by the North American Industry Classification System (NAICS) codes 11 through 99, are represented, except for the following: NAICS 111, 112, 4811 (part), 482, 491, 525 (part), 813, 814, and 92. 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. Data are presented by industry classifications and/or geographic area (states, metropolitan and micropolitan statistical areas, counties, and corporate municipalities (places) including cities, towns, townships, villages, and boroughs), and size of firm (employment and receipts).

Type of Data Collection Operation: The survey is based on a stratified probability sample of approximately 2.3 million firms from a universe of approximately 23 million firms. There were 5.5 million firms with paid employees and 17.4 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: Sampling error is present in these estimates because they are based on the results of a sample survey and not on an enumeration of the entire universe. Since these estimates are based on a probability sample, it is possible to estimate the sampling variability of the survey estimates. The standard error (SE) provides a measure of the variation. The relative SE or CV provides a measure of the magnitude of the variation relative to the estimate and is calculated as 100 multiplied by the ratio of the estimate to the SE. The CVs for number of firms and receipts at the national level typically range from 0 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. Explicit measures of the effects of these nonsampling errors are not available. However, it is believed that most of the important operational and data errors were detected and corrected through an automated data edit designed to review the data for reasonableness and consistency. Quality control techniques were used to verify that operating procedures were carried out as specified.

Sources of Additional Materials: U.S. Census Bureau, Guide to the 2002 Economic Census and Related Statistics.

U.S. DEPARTMENT OF EDUCATION
National Center for Education 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: Data are collected through a Web-based survey in the fall of every year. Missing data are imputed by using data of similar institutions.

Estimates of Sampling Error: Not applicable.

Other (nonsampling Errors: For 2004–05, approximately 99.9 percent response rate for degree-granting institutions.

Sources of Additional Material: U.S. Department of Education, National Center for Education Statistics, *Postsecondary Institutions in the United States: Fall 2005 and Degrees and Other Awards Conferred: 2004–05*. See <http://www.nces.ed.gov/ipeds/>.

National Household Education Surveys (NHES) Program

Universe, Frequency, and Types of Data: The National Household Education Surveys Program is a system of telephone surveys of the noninstitutionalized civilian population of the United States. Surveys in NHES have varying universes of interest depending on the particular survey. Specific topics covered by each survey are at the NHES Web site <http://nces.ed.gov/nhes>. A list of the surveys fielded as part of NHES, each universe, and the years they were fielded is provided below. 1) Adult Education—Interviews were conducted with a representative sample of civilian, noninstitutionalized persons age 16 and older who were not enrolled in grade 12 or below (1991, 1995, 1999, 2001, 2003, 2005). 2) After-School Programs and Activities—Interviews were conducted with parents of a representative sample of students in grades K through 8 (1999, 2001, 2005). 3) Civic Involvement—Interviews were conducted with representative samples of parents, youth, and adults (1996, 1999). 4) Early Childhood Program Participation—Interviews were conducted with parents of a representative sample of children from birth through grade 3, with the specific age groups varying by survey year (1991, 1995, 1999, 2001, 2005). 5) Household and Library Use—Interviews were conducted with a representative sample of U.S. households (1996). 6) Parent and Family Involvement in Education—Interviews were conducted with parents of a representative sample of children age three through grade 12 or in grades K through 12 depending on the survey year (1996, 1999, 2003, and 2007 forthcoming). 7) School Readiness—Interviews were conducted with parents of a representative sample of 3- to 7-year old children (1993, 1999, and 2007 forthcoming). 8) School Safety and Discipline—Interviews were conducted with a representative sample of students in grades 6–12, their parents, and the parents of a representative sample of students in grades 3–12 (1993).

Type of Data Collection Operation: NHES uses telephone interviews to collect data.

Data Collection and Imputation Procedures: Telephone numbers are selected using random digit dialing (RDD) techniques. Approximately 45,000 to 64,000 households are contacted in order to identify persons eligible for the surveys. Data are collected using computer-assisted telephone interviewing (CATI) procedures. Missing data are imputed using hot-deck imputation procedures.

Estimates of Sampling Error: Unweighted sample sizes range between 2,250 and 55,708. The average root design effects of the surveys in NHES range from 1.1 to 1.5, except for the Adult Education survey of 1991. In 1991, average root design effects for the Adult Education survey ranged from 2.3 to 4.5.

Other (nonsampling) Errors: Because of unit nonresponse and because the samples are drawn from households with telephone instead of all households, nonresponse and/or coverage bias may exist for some estimates. However, both sources of potential bias are adjusted for in the weighting process. Analyses of both potential sources of bias in the NHES collections have been studied and no significant bias has been detected.

Sources of Additional Material: Please see the NHES Web site at <http://nces.ed.gov/nhes>.

Schools and Staffing Survey (SASS)

Universe, frequency, and Types of Data: NCES designed the SASS survey system to emphasize teacher demand and shortage, teacher and administrator characteristics, school programs, and general conditions in schools. SASS also collects data on many other topics, including principals' and teachers' perceptions of school climate and problems in their schools; teacher compensation;

district hiring practices and basic characteristics of the student population. The SASS has had four core components: the School Questionnaire, the Teacher Questionnaire, the Principal Questionnaire, and the School District Questionnaire. For the 2003–04 SASS, a sample of public charter schools is included in the sample as part of the public school questionnaire. Since 1987–88, the SASS is the largest, most extensive survey of K through 12 school districts, schools, teachers, and administrators in the U.S. Surveys have been conducted every 3 to 4 years depending on budgetary constraints. The SASS includes data from public, private, and Bureau of Indian Affairs school sectors. Therefore, the SASS provides a multitude of opportunities for analysis and reporting on elementary and secondary educational issues

Type of Data Collection Operation: The U.S. Census Bureau performs the data collection and begins by sending advance letters to the sampled Local Education Agencies (LEAs) and schools in August and September of collection years. Beginning in October, questionnaires are delivered by U.S. Census Bureau field representatives. The sampling frame for the public school sample is the most recent Common Core of Data (CCD) school file. CCD is a universe file that includes all elementary and secondary schools in the United States. Schools operated by the Department of Defense or those that offered only kindergarten or pre-kindergarten or adult education were excluded from the SASS sample. The list frame used for the private school sample is the most recent Private School Universe Survey (PSS) list, updated with association lists. An area frame supplement is based on the canvassing of private schools within specific geographical areas. A separate universe of schools funded by the Bureau of Indian Affairs (BIA) is drawn from the Program Education Directory maintained by the BIA. To avoid duplicates in the BIA files, BIA schools in the CCD school file are treated as public schools.

Estimates of Sampling Error: Sample errors can be calculated using replicate weights and Balanced Repeated Replication complex survey design methodology. Errors depend on cell sizes and range from less than 1 percent to over 5 percent (for reasonable cell sizes).

Other (nonsampling) Errors: Because of unit nonresponse, bias may exist in some sample cells. However, bias has been adjusted for in the weighting process. Analysis of bias has been studied and no significant bias has been detected.

Sources of Additional Material: Please see the SASS web site at <http://nces.ed.gov/surveys/sass/>.

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 also collected on crimes cleared by arrest or exceptional means; by age, sex, and race of arrestees and for victims and offenders for homicides, number of law enforcement employees, on fatal and nonfatal assaults against law enforcement officers, and on hate crimes reported.

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: The coverage is 94 percent of the population (95 percent in MSAs, 87 percent in "cities outside of metropolitan areas," and 88 percent in nonmetropolitan counties) by UCR Program, through varying number of agencies reporting.

Sources of Additional Material: U.S. Federal Bureau of Investigation, *Crime in the United States, annual, Hate Crime Statistics*, annual, *Law Enforcement Officers Killed & Assaulted*, annual, <http://www.fbi.gov/ucr/ucr.htm>.

U.S. INTERNAL REVENUE SERVICE

Corporation Income Tax Returns

Universe, Frequency, and Types of Data: Annual study of unaudited corporation income tax returns, Forms 1120, 1120-A, 1120-F, 1120-L, 1120-PC, 1120-REIT, 1120-RIC, and 1120S, 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 146,000 returns for Tax Year 2004, allocated to sample classes which are based on type of return, size of total assets, size of net income or deficit, and selected business activity. Sampling rates for sample classes 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 comply with statistical definitions.

Estimates of Sampling Error: Estimated CVs for Tax Year 2004: Returns with assets over $10 million are self-representing. Coefficients of variation are published in the 2004 Statistics of Income Corporation Income Tax Returns, Table 1, by industry group.

Other (nonsampling) Errors: Nonsampling errors include coverage errors, processing errors, and response errors.

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

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 the population of returns filed, also include returns from 1040NR, filed by nonresident aliens plus certain self employment tax returns.

Type of Data Collection Operation: Annual 2004 stratified probability sample of approximately 201,000 returns broken into sample strata based on the larger of total income or total loss amounts, the size of business plus farm receipts, and other criteria such as the potential usefulness of the return for tax policy modeling. 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 2004: Adjusted gross income less deficit 0.11 percent; salaries and wages 0.21 percent; and tax exempt interest received 1.76 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, (Publication 1304).

Partnership Income Tax Returns

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 44,000 partnership returns from a population of 2.9 million filed during calendar year 2006 The sample is classified based on combinations of gross

receipts, net income or loss, total assets, and on industry. Sampling rates vary from 0.09 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 2005 (latest available): For number of partnerships, 0.32 percent; business receipts, 0.18 percent; Net income, 0.57 percent; net loss; 1.31 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. 26, No. 2 (Fall 2006).

Sole Proprietorship Income Tax Returns

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 63,000 sole proprietorships for tax year 2004. The sample is classified based on presence or absence of certain business schedules; the larger of total income or loss; size of business plus farm receipts, and other criteria such as the potential usefulness of the return for tax policy modeling. 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 2004 are available. For sole proprietorships, business receipts, 0.64 percent; depreciation 1.35 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 1980 through 1983) and *Statistics of Income Bulletin*, Vol. 26, No. 1 (Summer 2006, as well as bulletins for earlier years).

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

National Health Interview Survey (NHIS)

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

Type of Data Collection Operation: Multistage probability sample of 49,000 households (in 198 PSUs) from 1985 to 1994; 36–40,000 households (358 design PSUs or 449 effective PSUs when divided by state boundaries) from 1995 to 2005; an estimated completed 35,000 households (428 effective PSUs) beginning in 2006.

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

Estimates of Sampling Error: For 2004 medically attended injury episodes rates in the past 12 months by falling for: females 46.38 (3.50), and males 36.87 (3.24) per 1,000 population; for 2004 injury episodes rates during the past 12 months inside the home—29.72 (2.11) per 1,000 population.

Other (nonsampling) Errors: The response rate was 93.8 percent in 1996; in 2006, the total household response rate was 87.3 percent, with the final family response rate of 87.0 percent, and the final sample adult response rate of 70.8

percent. (Note: the NHIS questionnaire was redesigned in 1997, and a new sample design was instituted in 2006).

Sources of Additional Material: National Center for Health Statistics, Summary Health Statistics for the U.S. Population: National Health Interview Survey, 2005, Vital and Health Statistics, Series 10 #233; National Center for Health Statistics, Summary Health Statistics for U.S. Children: National Health Interview Survey, 2005, Vital and Health Statistics, Series 10 #231; National Center for Health Statistics, Summary Health Statistics for U.S. Adults: National Health Interview Survey, 2005, Vital and Health Statistics, Series 10 #232; 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.

National Survey of Family Growth (NSFG)

Universe, Frequency, and Types of Data: Periodic survey of men and women 15–44 years of age in the household population of the United States. Interviews were conducted in 2002 in person by trained female interviewers. Interview topics covered include births and pregnancies, marriage, divorce, and cohabitation, sexual activity, contraceptive use, and medical care. For men, data on father involvement with children were collected. The most sensitive data—on sexual behavior related to HIV and Sexually Transmitted Disease risk—were collected in a self-administered form in which the data are entered into a computer.

Type of Data Collection Operation: In the 2002 (Cycle 6) NSFG, the sample was a multistage area probability sample of men and women 15–44 years of age in the household population of the United States. Only one person 15–44 was selected from households with one or more persons 15–44. Data were collected and entered into laptop (notebook) computers. In the self-administered portion, the respondent entered his or her own answers into the computer. Sample included 12,571 interviews. The response rate was 79 percent. Hispanic and Black persons, as well as those 15–19 years of age, were sampled at higher rates than White adults. All percentages and other statistics shown for the NSFG are weighted to make national estimates. The weights adjust for the different rates of sampling for each group, and for nonresponse.

Data Collection and Imputation Procedures: When interviews are received, they are reviewed for consistency and quality, and analysis variables (recodes) are created. Missing data on these recodes were imputed using multiple regression techniques and checked again for consistency. Variables indicating whether a value has been imputed ("imputation flags") are included on the data file.

Estimates of Sampling Error: Sampling error codes are included on the data file so that users can estimate sampling errors for their own analyses. Sampling error estimates for nine illustrative analyses are shown on the NSFG Web site at <http://www.cdc.gov/nchs/nsfg.htm>. Sampling error estimates are also shown in most NCHS reports.

Other (nonsampling) Errors: In any survey, errors can occur because the respondent (the person being interviewed) does not recall the specific fact or event being asked about. The NSFG questionnaire in 2002 was programmed to check the consistency of many variables during the interview, so that the interviewer and respondent had a chance to correct any inconsistent information. Further checking occurred after the interview and during recoding and imputation. Typically, less than 1 percent of cases need imputation because of missing data.

Sources of Additional Material: The following references can be found at <http://www.cdc.gov/nchs/nsfg.htm>. "National Survey of Family Growth, Cycle 6: Sample Design, Weighting, and Variance Estimation." Vital and Health Statistics, Series 2, Number 142, July 2006. "Plan and Operation of Cycle 6 of the National Survey of Family Growth." Vital and Health Statistics, Series 1, No. 42. August 2005. "Sexual Behavior and Selected Health Measures: Men and Women 15–44 Years of Age, United States, 2002." *Advance Data from Vital and Health Statistics*, No. 362, Sept 15, 2005.

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 1951–1971, based on 50 percent sample of birth certificates, except a 20 percent to 50 percent sample 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 and are providing complete files of birth certificates.

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, there is no sampling for these files; the files are based on 100 percent of events registered.

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 the *National Vital Statistics Reports*. See the NCHS Web site at, <http://www.cdc.gov/nchs/nvss.htm>.

National Highway Traffic Safety Administration (NHTSA)

Fatality Analysis Reporting System (FARS)

Universe, Frequency, and Types of Data: FARS is a census of all fatal motor vehicle traffic crashes that occur throughout the United States including the District of Columbia and Puerto Rico on roadways customarily open to the public. The crash must be reported to the state/jurisdiction and at least one directly related fatality must occur within thirty days of the crash.

Type of Data Collection Operation: One or more analysts, in each state, extract data from the official documents and enter the data into a standardized electronic database.

Data Collection and Imputation Procedures: Detailed data describing the characteristics of the fatal crash, the vehicles and persons involved are obtained from police crash reports, driver and vehicle registration records, autopsy reports, highway department, etc. Computerized edit checks monitor the accuracy and completeness of the data. The FARS incorporates a sophisticated mathematical multiple imputation procedure to develop a probability distribution of missing blood alcohol concentration (BAC) levels in the database for drivers, pedestrians, and cyclists.

Estimates of Sampling Error: Since this is census data, there are no sampling errors.

Other (nonsampling) Errors: FARS represents a census of all police reported crashes and captures all data reported at the state level. FARS data undergo a rigorous quality control process to prevent inaccurate reporting. However, these data are highly dependent on the accuracy of the police accident reports. Errors or omissions within police accident reports may not be detected.

Sources of Additional Material: The FARS Coding and Validation Manual, ANSI D16.1 Manual on Classification of Motor Vehicle Traffic Accidents (Sixth Edition).