Chapter 1.
Labor Force Data Derived from the Current Population Survey

Each month, the Bureau of Labor Statistics (BLS) analyzes and publishes statistics on the labor force, employment, and unemployment, classified by a variety of demographic, social, and economic characteristics. These statistics are derived from the Current Population Survey (CPS), which is conducted by the Census Bureau for BLS. This monthly survey of the population uses a sample of households that is designed to represent the civilian noninstitutional population of the United States.

IN THIS CHAPTER

Background

Specific concepts of the labor force, employment, and unemployment were developed in the later stages of the Depression of the 1930s. Before the 1930s, aside from attempts in some of the decennial censuses, no direct measurements were made of the number of jobless persons. Mass unemployment in the early 1930s increased the need for statistics, and widely conflicting estimates based on a variety of indirect techniques began to appear. Dissatisfied with these methods, many research groups, as well as State and municipal governments, began experimenting with direct surveys or samples of the population. In these surveys, an attempt was made to classify the population as employed, unemployed, or out of the labor force by means of a series of questions addressed to each individual. In most of the surveys, the employed were defined as persons with occupations ("gainful workers"), and the unemployed were defined as those who were not working but were "willing and able to work." These concepts did not meet the standards of objectivity that many technicians felt were necessary to measure either the level of unemployment at a point in time or changes over time. Counts of gainful workers did not have a current dimension, and the criterion "willing and able to work," when applied in specific situations, appeared to be too intangible and too dependent upon the interpretation and attitude of the persons being interviewed.

A set of precise concepts was developed in the late 1930s to address these various criticisms. The classification of an individual depended principally upon his or her actual *activity* within a designated period, that is, was the individual working, looking for work, or engaged in other activities? These concepts were adopted for the national sample survey of households, called the Monthly Report of Unemployment, initiated in 1940 by the Works Progress Administration.

The household survey was transferred to the Census Bureau in late 1942, and its name was changed to the Monthly Report on the Labor Force. The name was changed once more, in 1948, to the present Current Population Survey in order to reflect the survey's expanding role as a source for data on a wide variety of demographic, social, and economic characteristics of the population. In 1959, responsibility for analyzing and publishing the CPS labor force data was transferred to BLS; the Census Bureau continues to collect the data.

Description of the Survey

The CPS collects information on the labor force status of the civilian noninstitutional population 15 years of age and older,

although labor force estimates are reported only for those 16 and older. Persons under 16 years of age are excluded from the official estimates because child labor laws, compulsory school attendance, and general social custom in the United States severely limit the types and amount of work that these children can do. Persons on active duty in the U.S. Armed Forces are excluded from coverage. The institutional population, which also is excluded from coverage, consists of residents of penal and mental institutions and homes for the aged and infirm.

The CPS is collected each month from a probability sample of approximately 60,000 households. Respondents are assured that all information obtained is completely confidential and is used only for the purpose of statistical analysis. Although the survey is conducted on a strictly voluntary basis, refusals to cooperate amount to only about 4 percent each month. (Another 3 to 4 percent of eligible households are not interviewed because of other failures to make contact.)

A calendar week was selected as the survey reference period because the period used must be short enough so that the data obtained are "current," but not so short that such occurrences as holidays or bad weather might cause erratic fluctuations in the information obtained. In addition, the reference period should not be so long that it challenges the recall of the respondent. A calendar week fulfills these conditions. Since July 1955, the calendar week, Sunday through Saturday, that includes the 12th day of the month has been defined as the reference week. The actual survey is conducted during the following week, the week containing the 19th day of the month.

Concepts

The criteria used in classifying persons on the basis of their labor force activity and some of the major statistics obtained from the CPS are as follows:

Employed persons. All those who, during the reference week, (1) did any work at all as paid employees, worked in their own business or profession or on their own farm, or worked 15 hours or more as unpaid workers in a family-operated enterprise; and (2) all those who did not work but had jobs or businesses from which they were temporarily absent due to illness, bad weather, vacation, childcare problems, labor dispute, maternity or paternity leave, or other family or personal obligations—whether or not they were paid by their employers for the time off and whether or not they were seeking other jobs. Each employed person is counted only once, even if he or she holds more than one job. Included in the total are employed citizens of foreign countries who are residing in the United States, but who are not living on the premises of an embassy. Excluded are persons whose only activity consisted of work around their own home (such as housework, painting, repairing, and so forth) or volunteer work for religious, charitable, and similar organizations.

Unemployed persons. All persons who: 1) had no employment during the reference week; 2) were available for work, except for temporary illness; and 3) had made specific efforts, such as contacting employers, to find employment sometime during the 4-week period ending with the reference week. Persons who were waiting to be recalled to a job from which they had been laid off need not have been looking for work to be classified as unemployed.

Duration of unemployment represents the length of time (through the current reference week) that persons classified as unemployed had been continuously looking for work. For persons on layoff, duration of unemployment represents the number of full weeks since the end of their most recent period of employment. Thus, it is a measure of an in-progress spell of joblessness, not a completed spell. Two useful measures of the duration of unemployment are the mean and the median. Mean duration is the arithmetic average computed from single weeks of unemployment. Median duration is the midpoint of a distribution of weeks of unemployment.

The reasons for unemployment are divided into four major groups: (1) Job losers, defined as (a) persons on temporary layoff, who have been given a date to return to work or who expect to return within 6 months (persons on layoff need not be looking for work to be classified as unemployed); (b) permanent job losers, whose employment ended involuntarily and who began looking for work; and (c) persons who completed a temporary job, and who began looking for work after the job ended; (2) Job leavers, defined as persons who quit or otherwise terminated their employment voluntarily and immediately began looking for work; (3) Reentrants, defined as persons who previously worked but were out of the labor force prior to beginning their job search; and (4) New entrants, defined as persons who never had worked but were searching for work.

Civilian labor force. This is the total of all civilians classified as employed and unemployed.

Unemployment rate. This represents the proportion of the civilian labor force that is unemployed.

Participation rate. This represents the proportion of the population that is in the labor force.

Employment-population ratio. This represents the proportion of the population that is employed.

Not in the labor force. Included in this group are all persons in the civilian noninstitutional population who are neither employed nor unemployed. Information is collected on their desire for and availability to take a job at the time of the CPS interview, jobsearch activity in the prior year, and reason for not looking for work in the 4-week period ending with the reference week. Persons not in the labor force who want and are available for a job and who have looked for work sometime in the past 12 months (or since the end of their last job if

they held one within the past 12 months), but who are not currently looking, are designated as "marginally attached to the labor force." The marginally attached are divided into those not currently looking because they believe their search would be futile—so-called discouraged workers—and those not currently looking for other reasons such as family responsibilities, ill health, or lack of transportation. For discouraged workers, the reasons for not currently looking for work are that the individual believes that: No work is available in his or her line of work or area; he or she could not find any work; he or she lacks necessary schooling, training, skills, or experience; employers would think he or she is too young or too old; or he or she would encounter hiring discrimination.

Multiple jobholders. These are employed persons who, during the reference week, had two or more jobs as a wage and salary worker, were self-employed and also held a wage and salary job, or worked as an unpaid family worker and also held a wage and salary job.

At work part time for economic reasons. Sometimes referred to as involuntary part time, this category refers to individuals who gave an economic reason for working 1 to 34 hours during the reference week. Economic reasons include: Slack work or unfavorable business conditions, inability to find full-time work, and seasonal declines in demand. Those who usually work part time must also indicate that they want and are available to work full time to be classified as part time for economic reasons.

At work part time for noneconomic reasons. This group includes those persons who usually work part time and were at work 1 to 34 hours during the reference week for a noneconomic reason. Noneconomic reasons include: Illness or other medical limitations, childcare problems or other family or personal obligations, school or training, retirement or Social Security limits on earnings, and being in a job in which full-time work is less than 35 hours. This group also includes those who gave an economic reason for usually working 1 to 34 hours but said they do not want to work full time or were unavailable for such work.

Usual full- or part-time status. Full-time workers are those who usually worked 35 hours or more (at all jobs combined). This group includes some individuals who worked less than 35 hours in the reference week for either economic or noneconomic reasons and those temporarily absent from work who usually work at least 35 hours per week. Part-time workers are those who usually work less than 35 hours per week (at all jobs), regardless of the number of hours worked in the reference week. This may include some individuals who actually worked more than 34 hours in the reference week, as well as those temporarily absent from work who usually work less than 35 hours.

Usual weekly earnings for wage and salary workers. Data are collected on earnings before taxes and other deductions, and include any overtime pay, commissions, or tips usually received (at the main job in the case of multiple jobholders). Earnings reported on a basis other than weekly (such as annual, monthly, or hourly) are converted to weekly. The term "usual" is as perceived by the respondent. If the respondent asks for a definition of usual, interviewers are instructed to define the term as more than half the weeks worked during the past 4 or 5 months.

Recent Changes to the Survey

Sample expansion

Beginning with the release of July 2001 data, labor force estimates from the CPS reflect the expansion of the monthly CPS sample from about 50,000 to about 60,000 eligible households. This expansion was one part of the Census Bureau's plan to meet the requirement of the State Children's Health Insurance Program (SCHIP) legislation. The SCHIP legislation requires the Census Bureau to improve State estimates of the number of children who live in low-income families and lack health insurance. These estimates are obtained from the Annual Demographic Supplement to the CPS (better known as the March income supplement).

In September 2000, the Census Bureau began expanding the monthly CPS sample in 31 States and the District of Columbia. The additional 10,000 households were added to the sample over a 3-month period. BLS chose not to include the additional households in the official labor force estimates, however, until it had sufficient time to evaluate the estimates from the expanded sample.

Estimates at the national level (not seasonally adjusted) derived from the 50,000- and 60,000-household samples were virtually the same. In any given month, the 60,000-household sample estimates for the overall labor force participation rate and the employment-population ratio differed by no more than 0.1 percentage point from estimates produced from the 50,000-household sample. The overall unemployment rates were identical in both samples. (For a discussion of the effect of the sample expansion on State estimates, see the forthcoming update of chapter 4.)

At the national level, previously published monthly labor force estimates for January to June 2001 were not revised, because the differences between the two samples were minimal. The 2001 annual averages for all labor force series, however, were calculated using the monthly average (January-December) from the expanded 60,000-household sample.

The 1994 redesign

A major redesign of the CPS was implemented in January 1994. The primary objective was to improve the quality of the data derived from the survey by introducing a new questionnaire and modernized data collection methods. Prior to 1994, the survey questionnaire had been virtually unchanged since 1967, at which time changes had been introduced based on recommendations of the Gordon Committee (President's

Committee to Appraise Employment and Unemployment Statistics, 1962). Additional changes were proposed in the late 1970s based on the recommendations of the Levitan Commission (National Commission on Employment and Unemployment Statistics, 1979); these, in part, formed the basis for the 1994 redesign.

The redesign of the questionnaire had four main objectives: 1) To adopt a computer-assisted interviewing environment, 2) to measure the official labor force concepts more precisely, 3) to expand the amount of data available, and 4) to implement several definitional changes.

Computerization. The new questionnaire was designed for a computer-assisted interview, in which interviewers ask the survey questions as they appear automatically on the screen of their laptop computer, and then type the responses directly into the laptop. In most cases, interviewers conduct the survey either in person at the respondent's home or by telephone from the interviewer's home. This mode of data collection is known as computer-assisted personal interviewing (CAPI). (In addition, about 10 percent of sample households are interviewed from centralized telephone centers, as explained below in the section on collection methods.)

Computer-assisted interviewing has important benefits, most notably that it facilitates the use of a relatively complex questionnaire that incorporates complicated skip patterns and standardized followup questions. Additionally, certain questions are automatically tailored to the individual's situation to make them more understandable. The computerized questionnaire also has several built-in editing features, including automatic checks for internal consistency and unlikely responses. An automated interview also permits dependent interviewing, that is, the use of information in the current interview that was obtained in a previous month's interview. Dependent interviewing reduces respondent and interviewer burden, while improving consistency of the data from one month to the next. The technique is being used to confirm the previously reported industry and occupation of a person's job, to calculate unemployment duration, and, for many people not in the labor force, to confirm their status as retired or disabled.

Major questionnaire changes. While the labor force status of most people is straightforward, some persons are more difficult to classify correctly, especially if they are engaged in activities that are relatively informal or intermittent. Many of the changes to the questionnaire were made to deal with such cases. This was accomplished by rewording and adding questions to conform more precisely to the official definitions, making the questions easier to understand and answer, minimizing reliance on volunteered responses, revising response categories, and taking advantage of the benefits of an automated interview. Areas affected by these improvements include:

1. *On layoff*. Persons on layoff are defined as those who are separated from a job to which they are awaiting recall. The old questionnaire, however, was not structured to consistently obtain information on the expectation of recall. In order to measure layoffs more accurately, questions were added to determine if people reported to be on layoff did in fact have an expectation of recall—that is, had they been given a specific date to return to work or, at least, had they been given an indication that they would be recalled within the next 6 months.

2. *Jobsearch methods.* To allow interviewers to better distinguish between active and passive methods, the response categories for jobsearch methods were expanded and reformatted. Also, the basic question on jobsearch methods was reworded and followup questions were added to encourage respondents to report all types of jobsearch activity.

3. *Hours at work*. To improve the accuracy of these data, the series of questions on hours worked was reordered to incorporate a recall strategy that asks for usual hours first, then about possible time taken off or extra hours worked during the reference week, and finally about hours actually worked.

4. *Reasons for working part time*. Persons who work part time do so either for noneconomic reasons (that is, because of personal constraints or preferences) or for economic reasons (that is, because of business-related constraints such as slack work or the lack of full-time opportunities). Because respondents typically are not familiar with this distinction, the question was reworded to provide examples of the two types of reasons. More importantly, the measurement of working part time involuntarily (or for economic reasons) was modified to better reflect the concept. Starting in 1994, workers who usually work part time and are working part time involuntarily must want and be available for full-time work.

5. *Earnings.* With the previous questionnaire, respondents were asked to report their earnings as a weekly amount, even though that may not have been the easiest way for them to recall or report their earnings. In the new version, respondents are asked to report earnings in the timeframe that they find easiest, for example, hourly, weekly, biweekly, monthly, or annual. Weekly earnings are automatically calculated for persons who respond on a basis other than weekly.

New data and definitional changes. The questionnaire redesign also made it possible to collect several types of data regularly for the first time, namely:

1. *Multiple jobholding*. Employed persons now are asked each month whether they had more than one job. This

allows BLS to produce estimates of multiple jobholding on a monthly basis, rather than having to derive them through special, periodic supplements.

2. Usual hours. All employed persons are asked each month about the hours they usually work. Previously, information on usual hours was collected from just one-quarter of wage and salary workers each month.

3. Other definitional changes. In addition, several labor force definitions were modified. The most important definitional changes concerned discouraged workers. The Levitan Commission had criticized the former definition because it was based on a subjective desire for work and on somewhat arbitrary assumptions about an individual's availability to take a job. As a result of the redesign, two requirements were added: For persons to qualify as discouraged, they must have engaged in some jobsearch within the past year (or since they last worked, if they worked within the past year), and they must be currently available to take a job. (Formerly, availability was inferred from responses to other questions; now, there is a direct question.) Also, beginning in January 1994, questions on this subject are asked of the full CPS sample, permitting estimates of the number of discouraged workers to be published monthly (rather than quarterly).

Another important definitional change concerned unemployed persons who were not working just before their jobsearch commenced, that is, new entrants or reentrants to the labor force. Prior to 1994, new entrants were defined as jobseekers who had never worked at a full-time job lasting 2 weeks or longer; reentrants were defined as jobseekers who had held a full-time job for at least 2 weeks and then had spent some time out of the labor force prior to their most recent period of jobsearch. These definitions were modified to encompass any type of job, not just a full-time job of at least 2 weeks' duration. Thus, new entrants now are defined as jobseekers who have never worked at all, and reentrants are jobseekers who have worked before, but not immediately prior to their jobsearch.

Changes Introduced in 2003

Several important changes were introduced into the survey in 2003. (For detailed information about these and other changes, see Bowler and others, 2003.)

New industrial and occupational classification systems

Information on the industry and occupation of the employed and unemployed is produced regularly from the CPS. The systems used to classify both industry and occupation were changed beginning with data published for January 2003. The 1990 Census Industrial Classification System was replaced by one based on the North American Industry Classification System (NAICS). Occupational data are being collected using new classifications derived from the Standard Occupational Classification (SOC) in lieu of the 1990 Census Occupational Classification System.

Population controls based on the 2000 census

New population controls based on the 2000 decennial census were introduced in the CPS beginning with data for January 2003. The new controls were prepared by projecting forward the civilian noninstitutional population as enumerated on April 1, 2000, and are used for the age-sex-race-ethnicity groups in the second-stage estimation procedure, as discussed in the section on estimation below.

New race and ethnicity categories

As a result of a directive issued by the U.S. Office of Management and Budget (OMB), all government statistics on race and ethnicity, including those from the CPS, are undergoing changes. Probably the most notable change is that survey respondents are given the opportunity to report themselves in more than one racial category. The racial categories are: White; black or African American, Asian, American Indian or Alaska Native; and Native Hawaiian or Other Pacific Islander. The questions used to obtain race and ethnicity in the CPS were modified to reflect the new directive, and publication tables were revised as well. Due to the limitations of the sample size, as well as the lack of population controls for the smaller race groups, data will be displayed for whites (no other race), blacks or African Americans (no other race), and Asians (no other race) only. In addition, the survey will continue to collect data on persons of Hispanic or Latino ethnicity. A direct question now is asked prior to the race questions to identify individuals as Spanish, Hispanic, or Latino; prior to 2003, the ethnicity of these persons was inferred from their country of origin. Under the OMB directive, Hispanic is still considered an ethnic, rather than a racial, category.

Redesigned CPS weighting

Modifications to basic and composite weighting procedures for the CPS also were implemented in January 2003. These changes were made based on a number of factors, including: BLS plans for publishing revised race categories at the State and national levels; making control-cell definitions more consistent across the second-stage weighting steps (State, ethnicity, and race), and between second-stage and composite weighting; precollapsing small cells to eliminate the "on-the-fly" collapsing algorithm that produces inconsistent results over time; and providing more stable monthly estimates for population subgroups of interest to users, including demographic population controls within each State.

Sampling

Since the inception of the survey, there have been various changes in the design of the CPS sample. The sample is traditionally redesigned and a new sample is selected after each decennial census. Also, the number of sample areas and the number of sample persons are changed occasion-

ally. Most of these changes are made in order to improve the efficiency of the sample design, increase the reliability of the sample estimates, or control costs. Since the mid-1980s, the CPS has had a State-based sample design, meaning that all sampling operations such as allocation and selection are implemented at the State level.

A redesigned CPS sample based on the 1990 decennial census was selected for use during the 1990s and the early years of the new century. Households from this new sample were phased into the CPS between April 1994 and July 1995. The July 1995 sample was the first monthly sample based entirely on the 1990 census. A redesigned sample based on the results of the 2000 census will be phased in between April 2004 and July 2005.

The original 1990 census-based sample design included about 66,000 housing units per month located in 792 selected geographic areas called primary sampling units (PSUs). The sample initially was selected to meet specific reliability criteria for the Nation, for each of the 50 States and the District of Columbia, and for the substate areas of New York City and the Los Angeles-Long Beach metropolitan area. In 1996, the original reliability criteria for the sample design were modified to reduce costs, which decreased the sample to 754 PSUs and 59,000 housing units. The current criteria, given below, are based on the coefficient of variation (CV) of the unemployment level, where the CV is defined as the standard error of the estimate divided by the estimate, expressed as a percentage. These CV controls assume a 6-percent unemployment rate in order to establish a consistent specification of sampling error.

The current sample design, including an expansion to meet the requirements of the SCHIP legislation, was introduced in July 2001. It includes about 72,000 households from 754 sample areas, or PSUs, and maintains a 1.9-percent CV on national monthly estimates of unemployment level. This translates into a change of 0.2 percentage point in the unemployment rate being significant at a 90-percent confidence level. For each of the 50 States and for the District of Columbia, the design maintains a CV of at most 8 percent on the annual average estimate of unemployment level, assuming a 6-percent unemployment rate. Due to the national reliability criterion, estimates for several large States are substantially more reliable than the State design criterion requires. Annual average unemployment estimates for California, Florida, New York, and Texas, for example, carry a CV of less than 4 percent.

In the first stage of sampling, the 754 PSUs are chosen. In the second stage, ultimate sampling unit clusters composed of about four housing units each are selected. Each month, about 72,000 housing units are assigned for data collection, of which about 60,000 are occupied and thus eligible for interview. The remainder are units found to be destroyed, vacant, converted to nonresidential use, containing persons whose usual place of residence is elsewhere, or ineligible for other reasons. Of the 60,000 housing units, about 7 to 8 percent are not interviewed in a given month due to temporary absence (vacation, for example) of the occupants, other failures to make contact after repeated attempts, inability of persons contacted to respond, unavailability for other reasons, and refusals to cooperate (about half of the noninterviews). Information is obtained each month for about 110,000 persons 16 years of age or older.

Selection of sample areas. The entire area of the United States, consisting of 3,141 counties and independent cities, is divided into 2,007 PSUs. PSUs are defined within States and do not cross State boundaries. In most States, a PSU consists of a county or a number of contiguous counties. In New England and Hawaii, minor civil divisions are used instead of counties.

Metropolitan areas within a State are used as a basis for forming many PSUs. Outside of metropolitan areas, two or more counties normally are combined to form a PSU except when the geographic area of an individual county is too large. Combining counties to form PSUs provides greater heterogeneity; a typical PSU includes urban and rural residents of both high and low economic levels and encompasses, to the extent feasible, diverse occupations and industries. Another important consideration is that the PSU be sufficiently compact so that, with a small sample spread throughout, it can be efficiently canvassed without undue travel costs.

The 2,007 PSUs are grouped into strata within each State. Then, one PSU is selected from each stratum with the probability of selection proportional to the population of the PSU. Nationally, there are a total of 428 PSUs in strata by themselves. These strata are self-representing and generally are the most populous PSUs in each State. The 326 remaining strata are formed by combining PSUs that are similar in such characteristics as unemployment, proportion of housing units with three or more persons, number of persons employed in various industries, and average monthly wages for various industries. The single PSU randomly selected from each of these strata is non-self-representing because it represents not only itself but the entire stratum. The probability of selecting a particular PSU in a non-self-representing stratum is proportional to its 1990 population. For example, within a stratum, the chance that a PSU with a population of 50,000 would be selected for the sample is twice that for a PSU having a population of 25,000.

Selection of sample households. Because the sample design is State based, the sampling ratio differs by State and depends on State population size as well as both national and State reliability requirements. The State sampling ratios range roughly from 1 in every 200 households to 1 in every 3,000 households. The sampling ratio occasionally is modified slightly to hold the size of the sample relatively constant given the overall growth of the population (called "sample maintenance reduction"). The sampling ratio used within a sample PSU depends on the probability of selection of the PSU and the sampling ratio for the State. In a sample PSU

with a probability of selection of 1 in 10 and a State sampling ratio of 3,000, a within-PSU sampling ratio of 1 in 300 achieves the desired overall ratio of 1 in 3,000 for the stratum.

The 1990 within-PSU sample design was developed using block-level data from the 1990 census. (The 1990 census was the first decennial census that produced data at the block level for the entire country.) Normally, census blocks are bounded by streets and other prominent physical features such as rivers or railroad tracks. County, Minor Civil Division, and census place limits also serve as block boundaries. In cities, blocks can be bounded by four streets and be quite small in land area. In rural areas, blocks can be several square miles in size.

For the purpose of sample selection, census blocks were grouped into three strata: Unit, group quarters, and area. (Occasionally, units within a block were split between the unit and group-quarters strata.) The unit stratum contained regular housing units with addresses that were easy to locate (for example, most single-family homes, townhouses, condominiums, apartment units, and mobile homes). The group-quarters stratum contained housing units in which residents shared common facilities or received formal or authorized care or custody. Unit and group-quarters blocks exist primarily in urban and suburban areas. The area stratum contains blocks with addresses that are more difficult to locate. Area blocks exist primarily in rural areas.

To reduce the variability of the survey estimates and to ensure that the within-PSU sample would reflect the demographic and socioeconomic characteristics of the PSU, blocks within the unit, group-quarters, and area strata were sorted using geographic and block-level data from the census. Examples of the census variables used for sorting include proportion of minority renter-occupied housing units, proportion of housing units with female householders, and proportion of owner-occupied housing units. The specific sorting variables used differed by type of PSU (urban or rural) and stratum.

Within each block, housing units were sorted geographically and grouped into clusters of approximately four units. A systematic sample of these clusters was then selected independently from each stratum using the appropriate within-PSU sampling ratio. The geographic clustering of the sample units reduces field representative travel costs. Prior to the interview, special listing procedures are used to locate the particular sample addresses in the group-quarters and area blocks.

Units in the three strata described above all existed at the time of the 1990 decennial census. Through a series of additional procedures, a sample of building permits is included in the CPS to represent housing units built after the decennial census. Adding these newly built units keeps the sample up-to-date and representative of the population. It also helps to keep the sample size stable: over the life of the sample, the addition of newly built housing units compensates for the loss of "old" units that may be abandoned, demolished, or converted to nonresidential use. In normal circumstances, the number of eligible households in the sample grows slowly. Sample maintenance reduction procedures are periodically implemented to hold the size of the sample relatively constant.

Rotation of sample. Part of the sample is changed each month. Each monthly sample is divided into eight representative subsamples or rotation groups. A given rotation group is interviewed for a total of 8 months, divided into two equal periods. The group is in the sample for 4 consecutive months, leaves the sample during the following 8 months, and then returns for another 4 consecutive months. In each monthly sample, 1 of the 8 rotation groups is in the first month of enumeration, another rotation group is in the second month, and so on. (The rotation group in the fifth month of enumeration is returning after an 8-month break.) Under this system, 75 percent of the sample is common from month to month and 50 percent is common from year to year for the same month. This procedure provides a substantial amount of month-to-month and year-to-year overlap in the sample, thus yielding better estimates of change and reducing discontinuities in the series of data without burdening sampled households with an unduly long period of inquiry.

Collection Methods

Each month, during the calendar week containing the 19th day, interviewers contact a "responsible" person in each of the sample households in the CPS. At the time of the first enumeration of a household, the interviewer visits the household and prepares a roster of the household members, including their personal characteristics (date of birth, sex, race, ethnic origin, marital status, educational attainment, veteran status, and so on) and their relationship to the person maintaining the household. The interviewers enter this information into laptop computers. This roster is then checked for accuracy and brought up to date at each subsequent interview to take account of new or departed residents, changes in marital status, and similar items. The information on personal characteristics is thus available each month for identification purposes and for cross-classification with economic characteristics of the sample population.

Personal visits are preferred in the first month in which the household is in the sample. In other months, the interview generally is conducted by telephone. Approximately 70 percent of the households in any given month are interviewed by telephone. A portion of the households (10 percent) is interviewed via computer-assisted telephone interviewing (CATI), from three centralized telephone centers (located in Hagerstown, MD; Jeffersonville, IN; and Tucson, AZ) by interviewers who also use a computerized questionnaire.

At each monthly visit, a series of standard questions on labor market activity during the preceding week is asked about each household member 15 years of age and older. (As previously mentioned, the official labor force estimates per-

tain to those aged 16 and older.) The primary purpose of these questions is to classify the sample population into the three basic economic groups: The employed, the unemployed, and those not in the labor force.

At the end of each day's interviewing, the data collected are transmitted to the Census Bureau's central computer in Washington, DC. Once files are transmitted to the main computer, they are deleted from the laptops.

Because of the crucial role interviewers have in the household survey, a great amount of time and effort is spent maintaining the quality of their work. Interviewers are given intensive training, including classroom lectures, discussion, practice, observation, home-study materials, and on-the-job training. At least once a year, they convene for daylong training and review sessions, and, also at least once a year, they are accompanied by a supervisor during a full day of interviewing to determine how well they carry out their assignments.

A selected number of households are reinterviewed each month to determine whether the information obtained in the first interview was correct. The information gained from these interviews is used to improve the entire training program.

Estimation Methods

Under the estimating methods used in the CPS, all of the results for a given month become available simultaneously and are based on returns from all respondents. The estimation procedure involves weighting the data from each sample person by the inverse of the probability of the person being in the sample. This gives a rough measure of the number of actual persons that the sample person represents. Since 1985, most sample persons within the same State have had the same probability of selection.

Some selection probabilities may differ within a State due to the sample design or for operational reasons. Field subsampling, for example, which is carried out when areas selected for the sample are found to contain many more households than expected, may cause probabilities of selection to differ for some sample areas within a State. Through a series of estimation steps (outlined below), the selection probabilities are adjusted for noninterviews and survey undercoverage; data from previous months are incorporated into the estimates through the composite estimation procedure.

1. *Noninterview adjustment*. The weights for all interviewed households are adjusted to account for occupied sample households for which no information was obtained because of the occupants' absence, impassable roads, refusals, or unavailability of the respondents for other reasons. This noninterview adjustment is made separately for clusters of similar sample areas that are usually, but not necessarily, contained within a State. Similarity of sample areas is based on Metropolitan Statistical Area (MSA) status and size. Within each cluster, there is a further breakdown by residence. Each MSA cluster is split into "central city" and "balance of the MSA." Each non-MSA cluster is split into "urban" and "rural" residence categories. The proportion of sample households not interviewed varies from 7 to 8 percent, depending on weather, vacation times, and so forth.

2. *Ratio estimates*. The distribution of the population selected for the sample differs by chance from that of the population as a whole in such characteristics as age, race, sex, ethnicity, and State of residence. Because these characteristics are closely correlated with labor force participation and other principal measurements made from the sample, the survey estimates can be substantially improved when weighted appropriately by the known distribution of these population characteristics. This is accomplished through two stages of ratio adjustment, as follows:

   a. *First-stage ratio estimation*. The purpose of the first-stage ratio adjustment is to reduce the contribution to variance of selecting a sample of PSUs rather than drawing sample households from every PSU in the Nation. This adjustment is made to the CPS weights in two race cells: Black and nonblack; and two age cells: 0 to 15 years and 16 years and older; it is applied only to data from PSUs that are not self-representing and for those States that have a substantial number of black households. The procedure corrects for differences that existed in each State cell at the time of the 2000 census between 1) the race distribution of the population in sample PSUs and 2) the race distribution of all PSUs. (Both 1 and 2 exclude self-representing PSUs.)

   b. *Second-stage ratio estimation*. This procedure substantially reduces the variability of estimates and corrects, to some extent, for CPS undercoverage. The CPS sample weights are adjusted to ensure that sample-based estimates of population match independent population controls.

   Beginning in 2003, the second-stage weighting has new coverage steps "0A" and "0B" that are followed by an iterative raking process. California and New York are split into substate areas, and 53 States/areas are used in Step 0B and Step 1 (Los Angeles-Long Beach metropolitan area; balance of California; New York City; balance of New York; the other 48 States; and the District of Columbia.)

   The noniterated National Coverage Step 0A is added primarily to improve the efficiency of adjustment for subpopulations that are prone to undercoverage. Step 0A also provides some control for Asian race that could not be included in the iterated steps.

   The noniterated State Coverage Step 0B is designed

to adjust for race/gender/age coverage differences between the States. Race is limited to black and nonblack, and there is no ethnicity component in the step.

The three iterated steps adjust sample weights to the following control groups:

i. State step—6 gender x age cells defined for 53 States/areas
ii. Ethnicity step—26 Hispanic and 26 non-Hispanic gender x age cells
iii. Race step—34 white-only, 26 black-only, and 26 Asian-only and residual gender x age cells

The independent population controls are prepared by projecting forward the resident population as enumerated on April 1, 2000. The projections are derived by updating demographic census data with information from a variety of other data sources that account for births, deaths, and net migration. Subtracting estimated numbers of resident Armed Forces personnel and institutionalized persons reduces the resident population to the civilian noninstitutional population.

3. *Composite weighting procedure*. The last step in the preparation of most CPS estimates makes use of a composite estimation procedure. Composite estimates are created as a weighted average of two factors: (1) The two-stage ratio estimate based on data from the entire sample for the current month; and (2), the composite estimate for the previous month, adjusted by an estimate of the month-to-month change based on the six rotation groups common to both months. A bias adjustment term is added to the weighted average to reduce variance and partially account for bias associated with month-in-sample estimates. This month-in-sample bias is exhibited by unemployment estimates for persons in their first and fifth months in the CPS that are generally higher than estimates obtained for the other months.

These composite estimates are then used as controls in the composite weighting procedure. Both employment and unemployment are controlled in each defined cell, and not-in-labor-force (NILF) is controlled as a residual. The iterative procedure is similar to that used for second-stage weighting:

a. State step—a single CPS16+ cell is used for 53 States/areas

b. Ethnicity step—10 Hispanic and 10 non-Hispanic gender x age cells

c. Race step—22 white-only, 14 black-only, and 10 Asian-only and residual gender x age cells

Composite estimation results in a reduction in the sampling error beyond that which is achieved through the two stages of ratio estimation. For some items, the reduction is substantial. The resultant gains in reliability are greatest in estimates of month-to-month change, although gains also are usually obtained for estimates of level in a given month, change from year to year, and change over other intervals of time.

Seasonal Adjustment

Over the course of a year, the size of the Nation's labor force, the levels of employment and unemployment, and other measures of labor market activity undergo sharp fluctuations due to such seasonal events as changes in weather, reduced or expanded production, harvests, major holidays, and the opening and closing of schools. Because these seasonal events follow a more or less regular pattern each year, their influence on statistical trends can be eliminated by adjusting the statistics from month to month. These adjustments make it easier to observe the cyclical and other nonseasonal movements in the series. In evaluating changes in a seasonally adjusted series, it is important to note that seasonal adjustment is merely an approximation based on past experience. Seasonally adjusted estimates have a broader margin of possible error than do the original data on which they are based, because they not only are subject to sampling and other errors but also are affected by the uncertainties of the seasonal adjustment process itself.

Since January 1980, national labor force data were seasonally adjusted with a procedure called *X-11 ARIMA* (Auto-Regressive Integrated Moving Average), which was developed at Statistics Canada as an extension of the standard X-11 method.

Beginning with data published for January 2003, the X-12-ARIMA program is used to seasonally adjust labor force series. This program includes all the capabilities of the Statistics Canada program and it introduces new features. Developed at the U.S. Census Bureau (Findley and others, 1988), X-12-ARIMA provides enhancements to (1) ARIMA time series modeling and model selection, (2) detection and estimation of outlier, trading day, and holiday effects, (3) post-adjustment diagnostics, and (4) seasonal and trend filter options.

The changes introduced into the CPS in 2003 affect the number of series that are directly seasonally adjusted. Prior to 2003, 182 series based on age, sex, industry, occupation, and other characteristics were directly seasonally adjusted; beginning in 2003, 116 series are directly seasonally adjusted. Eighty-one series were eliminated; most of these were related to industry and occupation. Fifteen aggregate series, previously derived from detailed series no longer seasonally adjusted, are now directly seasonally adjusted at the aggregate level.

At the beginning of each calendar year, projected seasonal adjustment factors are calculated for use during the January-June period. In July of each year, BLS calculates and publishes in *Employment and Earnings* projected seasonal adjustment factors for use in the second half of the year, based on the experience through June. Revisions of

historical data, usually for the most recent 5 years, are made only at the beginning of each calendar year.

Presentation and Uses

The CPS provides comprehensive information on the social, demographic, and economic characteristics of the civilian noninstitutional population 16 years of age and older.

Each month, the employment and unemployment data are published initially in The Employment Situation news release about 2 weeks after they are collected. The release includes a narrative summary and analysis of the major employment and unemployment developments, together with tables containing statistics for the principal data series. The news release is available on the Internet and can be accessed via the World Wide Web. The Universal Resource Locator is: **http://www.bls.gov/cps/home.htm**. The news release also is available on the BLS fax-on-demand service.

More detailed statistics are subsequently published in *Employment and Earnings*, a monthly periodical. The detailed tables provide information on the labor force, employment, and unemployment by a number of characteristics, such as age, sex, race, marital status, industry, and occupation. In addition, the January issue of *Employment and Earnings* provides annual averages for employment and earnings by detailed occupational categories and union affiliation, as well as estimates of employee absences.

Thousands of labor force data series are maintained in LABSTAT, the BLS public database on the Internet. They can be accessed at **http://www.bls.gov/cps/cpsatabs.htm**, **http://data.bls.gov/cgi-bin/surveymost?ln**, and **http://www.bls.gov/data/home.htm.**

The CPS also is used to obtain detailed information on particular segments of the population and labor force. Generally, these "supplemental" inquiries are repeated annually or biennially in the same month and include topics such as annual earnings and total incomes of individuals and families (published by the Census Bureau); the extent of work experience of the population during the prior calendar year; the employment of school-age youths, high school graduates, and dropouts; contingent workers; job tenure; displaced workers; and disabled veterans. Some additional supplements that are unrelated to labor force issues, such as those on smoking and voting, also are conducted through the CPS. Supplemental questions are asked following the completion of the regular monthly labor force questions.

Generally, the persons who provide information for the monthly CPS questions also answer the supplemental questions. Occasionally, the kind of information sought in the special survey requires the respondent to be the person about whom the questions are asked. Results of these special surveys usually are published in news releases and in the *Monthly Labor Review* or BLS reports.

In addition to the regularly tabulated statistics described above, special data can be generated from the public-use versions of CPS individual record (microdata) files. These files contain records of the responses to the survey questionnaire for all individuals in the survey, edited to protect the confidentiality of the respondents. While the files can be used simply to create additional cross-sectional detail, an important feature of their use is the ability to match the records of specific individuals at different points during their participation in the survey. By matching these records, data files can be created that lend themselves to some limited longitudinal analysis and the investigation of short-run labor market dynamics. Microdata files are available for all months since January 1976 and for various months in prior years. These data are made available on CD-ROM. Address inquiries regarding these files to: Division of Data Development and Publications, Bureau of Labor Statistics, Room 4965, 2 Massachusetts Ave., NE., Washington, DC 20212-0001, telephone 202-691-6345; e-mail *cpsinfo@bls.gov*.

Limitations of the Data

Geographic. Although the present CPS sample is a State-based design, the sample size of the CPS is sufficient to produce reliable monthly estimates at the national level only. The sample does not permit the production of reliable monthly estimates for the States. However, demographic, social, and economic detail is published annually for the census regions and divisions, all States and the District of Columbia, 50 large metropolitan areas, and selected central cities. The production of subnational labor force and unemployment estimates is discussed in more detail in chapter 4 of this bulletin.

Sources of errors in the survey estimates. There are two types of errors possible in an estimate based on a sample survey—sampling and nonsampling. The mathematical discipline of sampling theory provides methods for estimating standard errors when the probability of selection of each member of a population can be specified. The standard error, a measure of sampling variability, can be used to compute confidence intervals that indicate a range of differences from true population values that can be anticipated because only a sample of the population has been surveyed. Nonsampling errors such as response variability, response bias, and other types of bias occur in complete censuses as well as sample surveys. In some instances, nonsampling error may be more tightly controlled in a well-conducted survey, through which it is feasible to collect and process the data more skillfully. Estimation of other types of bias is one of the most difficult aspects of survey work, and adequate measures of bias often cannot be made.

Nonsampling error. The full extent of nonsampling error is unknown, but special studies have been conducted to quantify some sources of nonsampling error in the CPS. The effect of nonsampling error should be small on estimates of relative change, such as month-to-month change. Estimates

of monthly levels would be more severely affected by nonsampling error.

Nonsampling errors in surveys can be attributed to many sources, including the inability to obtain information about all persons in the sample; differences in the interpretation of questions; inability or unwillingness of respondents to provide correct information; inability to recall information; errors made in collecting and processing the data; errors made in estimating values for missing data; and failure to represent all sample households and all persons within sample households (undercoverage).

The effects of some components of nonsampling error in the CPS data are reflected in the variation in some labor force measures among the rotation groups, each of which is designed to be a representative sample of the population. For example, unemployment estimates from a rotation group tend to be higher in the first and fifth months of interviewing.

Undercoverage in the CPS results from missed housing units and missed persons within sample households. The noninterview adjustment procedure accounts for missed households. It also is known that the CPS undercoverage of persons varies with age, sex, race, and Hispanic ethnicity. Generally, undercoverage is greater for men than for women and greater for blacks, Hispanics, and other races than for whites. Ratio adjustment to independent age-sex-race-origin population controls, as described previously, partially corrects for the biases due to survey undercoverage. Biases still exist in the estimates to the extent that persons in missed households or missed persons in interviewed households have characteristics different from those of interviewed persons in the same age-sex-race-origin group.

The independent population estimates used in the estimation procedure may be a source of error, although, on balance, their use substantially improves the statistical reliability of many of the figures. Errors may arise in the independent population estimates because of underenumeration of certain population groups or errors in age reporting in the decennial census (which serves as the base for the estimates) or similar problems in the components of population change (mortality, immigration, and so forth) since that date.

Sampling error. When a sample, rather than the entire population, is surveyed, estimates differ from the true population values that they represent. This difference, or sampling error, occurs by chance, and its variability is measured by the standard error of the estimate. Sample estimates from a given survey design are unbiased when an average of the estimates from all possible samples would yield, hypothetically, the true population value. In this case, the sample estimate and its standard error can be used to construct approximate confidence intervals, or ranges of values, that include the true population value with known probabilities. If the process of selecting a sample from the population were repeated many times and an estimate and its standard error were calculated for each sample, then:

1. Approximately 68 percent of the intervals from 1 standard error below the estimate to 1 standard error above the estimate would include the true population value.
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 true population value.
3. Approximately 95 percent of the intervals from 2 standard errors below the estimate to 2 standard errors above the estimate would include the true population value.

Although the estimating methods used in the CPS do not produce unbiased estimates, biases for most estimates are believed to be small enough that these confidence interval statements are approximately true.

Standard error estimates computed using generalized variance functions are provided in *Employment and Earnings* and other publications. Using replicate variance techniques, standard error estimates are generated. As computed, these standard error estimates reflect contributions not only from sampling error, but also from some types of nonsampling error, particularly response variability. Because replicate variance techniques are somewhat cumbersome, simplified formulas called generalized variance functions (GVFs) have been developed for various types of labor force characteristics. The GVF can be used to approximate an estimate's standard error, but this indicates only the general magnitude of the standard error, rather than a precise value. ■

Technical References

Bowler, Mary, Randy E. Ilg, Stephen Miller, Anne Polivka, and Ed Robison. "Revisions to the Current Population Survey Effective in January 2003," *Employment and Earnings*, February 2003.

Bregger, John E. "The Current Population Survey: a historical perspective and BLS' role," *Monthly Labor Review*, June 1984, pp. 8-14.

Butz, W. P. and T.J. Plewes. "A Current Population Survey for the 21st Century," *Proceedings of the Fifth Annual Research Conference*, U.S. Department of Commerce, Bureau of the Census: Washington, DC, 1989, pp. 3-13.

Cohany, Sharon R., Anne E. Polivka, and Jennifer M. Rothgeb. "Revisions in the Current Population Survey Effective January 1994," *Employment and Earnings*, February 1994.

Dagum, Estela Bee. *The X-11 ARIMA Seasonal Adjustment Method*. Ottawa, Statistics Canada, January 1983 (Statistics Canada Catalogue No. 12-564E).

Findley, D. F., B.C. Monsell, W.R. Bell, M.C. Otto, and B.C. Chen. "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program," in *Journal of Business and Economic Statistics*, 1988, vol. 16, pp. 127-76 (with Discussion).

Levine, Chester, Laurie Salmon, and Daniel H. Weinberg. "Revising the Standard Occupational Classification System," *Monthly Labor Review*, May 1999, pp. 36-45.

National Commission on Employment and Unemployment Statistics. *Counting the Labor Force*, 1979.
A comprehensive review of the entire labor market data system; includes an appraisal of pre-1994 concepts and methodology and recommendations for further research and improvements.

"Overhauling the Current Population Survey," *Monthly Labor Review*, September 1993, pp. 3-33.
A trilogy of articles that provides additional information on the background of the CPS redesign.

President's Committee to Appraise Employment and Unemployment Statistics. *Measuring Employment and Unemployment*, 1962.
A review of all Federal statistical series on employment and unemployment and a comparison of the sources and uses of each series.

Tiller, Richard B. and Thomas Evans. "Revision of Seasonally Adjusted Labor Force Series," *Employment and Earnings*, February 2003.

U.S. Department of Commerce, Office of Federal Statistical Policy and Standards. "An Error Profile: Employment as Measured by the Current Population Survey," *Statistical Policy Working Paper* 3, 1978.
A description of the potential sources of error in the CPS as they affect the national employment statistics.

U.S. Department of Labor, Bureau of Labor Statistics. *Employment and Earnings*, "Explanatory Notes and Estimates of Error," monthly.
An up-to-date, concise description of the concepts and methods used in the labor force data from the Current Population Survey. Provides tables that present sampling errors for various labor force series.

U.S. Department of Labor, Bureau of Labor Statistics, and U.S Department of Commerce, U.S. Census Bureau. *Current Population Survey: Design and Methodology*, Technical Paper 63RV, 2002.
The most comprehensive guide to the survey. See also the Web site: **http://www.census.gov/prod/2002pubs/tp63rv.pdf**.

U.S. Department of Labor, Bureau of Labor Statistics. "Expansion of the Current Population Survey Sample Effective July 2001," *Employment and Earnings*, August 2001, pp. 3-7.

U.S. Department of Labor, Bureau of Labor Statistics. *Geographic Profile of Employment and Unemployment, 2000*, Bulletin 2550, 2002.
A tabular presentation of 2000 annual average data on the labor force characteristics of the population residing in the census regions, the 50 States and the District of Columbia, 50 large metropolitan areas, and 17 central cities.

U.S. Department of Labor, Bureau of Labor Statistics. *How the Government Measures Unemployment*, Report 864, 1994.
A short, nontechnical discussion of the concepts and methods used to obtain labor force statistics from the Current Population Survey. Updated version: **http://www.bls.gov/cps/cps_htgm.htm.**

Walker, James A. and John B. Murphy. "Implementing the North American Industry Classification System at BLS," *Monthly Labor Review*, December 2001, pp. 15-21.