Appendix B. SOURCE AND ACCURACY OF DATA

Source of Data

The data for this report, which cover a wide range of topics and years, came from the Current Population Survey (CPS), the Survey of Income and Program Participation (SIPP), the American Housing Survey (AHS), and the decennial censuses conducted by the Census Bureau. The surveys' estimation procedure adjusts weighted sample results to agree with independent estimates of the civilian noninstitutional population of the United States by age, sex, race, Hispanic/non-Hispanic ancestry, and state of residence.

The independent estimates are calculated based on information from four primary sources: the 1990 Decennial Census of Population and Housing, statistics on births, deaths, immigration, and emigration; statistics on the size of the Armed Forces; and starting in 1994, an adjustment for undercoverage in the 1990 decennial census. The estimation procedure for 1994 and later years used independent estimates based on the most recent decennial census at that time. (Data in some sections are revised for years prior to 1994.) This change in independent estimates had relatively little impact on summary measures, such as medians and percent distributions, but did have a significant impact on levels. For example, use of the 1990-based population controls resulted in about a 1-percent increase in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for 1994 and later years will differ from those for earlier years by more than what could be attributed to actual changes in the population. These differences could be disproportionately greater for certain population subgroups than for the total population. The estimation procedures for the CPS, SIPP, and AHS data are discussed in more detail in the publications cited in Appendix A of this report.

Reliability of Estimates

Since the CPS, SIPP, and AHS estimates come from samples, they may differ from the figures from a complete census using the same questionnaires, instructions, and enumerators. This possible variation in the estimates due to sampling is known as "sampling variability." A sample survey estimate has two types of error: sampling and nonsampling. The accuracy of an estimate depends on both types of error. The nature of the sampling error is known given the survey design. The full extent of nonsampling error, however, is unknown.

To estimate the standard error of a CPS estimate, the Census Bureau uses replicated variance estimation methods. These methods primarily measure the magnitude of sampling error. However, they do measure some effects of nonsampling error as well. They do not measure systematic biases in the data due to nonsampling error. (Bias is the average of the differences, over all possible samples, between the sample estimates and the desired value.)

Since the full extent of nonsampling error is unknown, one should be particularly careful when interpreting results based on small differences between the estimates. Even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test. Caution should also be used when interpreting results based on a relatively small number of cases. Summary measures probably do not reveal useful information when computed on a base smaller than 75,000.

Sampling Error

Standard errors are not given in this report because of the wide range of topics included and the wide variety of data sources. Standard errors may be found in the publications that are noted at the end of most sections and in Appendix A or by contacting the subject specialist provided at the end of each section.

Nonsampling Variability

As in any survey work, the results are subject to errors of response and nonreporting in addition to sampling variability. Nonsampling errors can be attributed to many sources, including:

- Inability to obtain information about all cases
- Definitional difficulties
- Differences in the respondent interpretation of questions
- Respondent inability or unwillingness to provide correct information
- Respondent inability to recall information
- Errors made in collection such as recording or coding data
- Errors made in processing the data
- Errors made in estimating values for missing data
- Failure to represent all units with the sample (undercoverage)

Comparability of Data

Data obtained from sample surveys and other sources are not entirely comparable. This results from differences in interviewer training and experience and in differing survey processes. This is an example of nonsampling variability not reflected in the standard errors. Therefore, caution should be used in comparing results from different sources.

A number of changes were made in data collection and estimation procedures beginning with the January 1994 CPS. The major change was the use of a new questionnaire. The questionnaire was redesigned to measure the official labor force concepts more precisely, to expand the amount of data available, to implement several definitional changes, and to adapt to a computer-assisted interviewing environment. The March supplemental income questions were also modified for adaptation to computer-assisted interviewing, although there were no changes in definitions and concepts. Because of these and other changes, caution should be used when comparing estimates from data collected before 1994 with estimates from data collected in 1994 or later. See the publications noted in Appendix A and at the end of most sections for a description of these changes and the effect they had on the data.

Census 2000 data were used in Chapter 2 of this report, while all other chapters in this report use survey data. This report includes data for three different population universes: resident population (Census 2000): civilian noninstitutional population, plus Armed Forces living off post or with their families on post (SIPP and March CPS universe), as well as the universe of housing units (AHS). Estimates using sample data from the CPS, SIPP, and the AHS are weighted by population controls based on the 1990 decennial census adjusted for estimated net undercount. As such, these estimates are not consistent with population estimates computed from either the intercensal estimates program (which are not adjusted for estimated net census undercount), or the 2000 decennial census.

Two different methods are being used to evaluate undercount in Census 2000: Demographic Analysis (DA) and the Accuracy and Coverage Evaluation (ACE). For more information on the accuracy of the 2000 decennial census, see the Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy, *(www.census.gov/dmd/www/pdf/Escap2.pdf).*