Second Quarter 2009 - Source and Accuracy of Estimates

Changes in 2007

In first quarter 2007, the CPS/HVS began using Blaise, a powerful computer-assisted interviewing (CAI) system and survey processing tool for the Windows operating system. It is being used for many of the surveys now being conducted by the Census Bureau.

Methodology Changes in 2003

Population controls that reflect the results of the 2000 decennial census were used in the CPS/HVS estimation process for the first time in the first quarter 2003. This change had a slight effect on vacancy and homeownership rates, as described below. As a final additional step in the estimation process, the estimates were controlled to independent housing counts used for the first time in order to produce a more accurate estimate of housing units. This should make the CPS/HVS estimates of housing units more consistent with other Census Bureau housing surveys. The new housing controls affected the count of all housing units in the sense that both occupied and vacant units were ratio estimated to the new control total. Vacancy rates and homeownership rates were not affected by this change.

The CPS/HVS began computing first-stage factors (used for weighting purposes) based on year-round and seasonal counts of housing units from the 2000 decennial census in the first quarter 2003. From 1980 to 2002, the CPS/HVS first-stage factors were based on year-round estimates only. We believe that this improves our counts of year-round and seasonal units.

The shift from 1990-based to 2000-based population controls (including the weighting revision) had a very slight effect on vacancy rates and homeownership rates. Research has shown that the new 2000-based controls dropped the rental vacancy rate in the first quarter 2002 from 9.14 percent to 9.08 percent---a difference of less than 1/10 of one percent. The homeowner vacancy rate was revised from 1.67 percent to 1.65 percent, while the homeownership rate was revised from 67.82 percent to 67.81 percent.

The questions on race on the CPS were modified beginning in the first quarter 2003 to comply with the revised standards for federal statistical agencies. Respondents may now select more than one race. The Hispanic/Nonhispanic origin question continues to be asked separately.

INTRODUCTION OF COMPUTER ASSISTED SURVEY COLLECTION (CASIC) AND REWEIGHTING CURRENT POPULATION SURVEY/HOUSING VACANCY SURVEY (CPS/HVS)–BASED ON THE 1990 CENSUS

Major changes related to the Current Population Survey/Housing Vacancy Survey (CPS/HVS) were effective beginning with the first quarter 1994 data. First, a new weighting procedure was implemented based on the 1990 decennial census. The 1990-based weighting produces, on average, estimates of the total housing inventory that were about 0.1 percent lower than the 1980-based weighting. Generally, the vacancy rates were only minimally affected, while the homeownership rate was about one-half of a percentage point lower with the new weighting procedures.

A second change was that the CPS/HVS became a totally computerized survey with the implementation of the Computer Assisted Survey Information Collection (CASIC). The CASIC tools consist of state-of-the-art computer-assisted modules for data collection and processing. Although the concepts, definitions, and questionnaire items remain the same, the shift to CASIC may affect vacancy rates and homeownership rates. We were unable to determine the quantitative effects of the use of CASIC on the vacancy and homeownership rates. Data users should use caution when comparing data for 1994 and later with earlier data.

Beginning the second quarter of 1999, a change was made in the way data for housing units in structure were collected. In the past, there was one category to show 1-unit in structure. That has been broken into two categories: 1-unit attached and 1-unit detached.

Source of Data

The estimates presented in this report were based on data obtained from two surveys conducted by the Bureau of the Census. Data concerning vacancy rates and tenure of occupied housing units were from the monthly sample of the Current Population Survey (CPS). Characteristics of occupied housing units in table 3 were from the American Housing Survey (AHS).

CPS and AHS Designs

Since the inception of the CPS in 1940, the sample has been redesigned several times to upgrade the quality and reliability of the data and to meet changing data needs. From July 1995 to March 2004, the CPS/HVS sample was selected from a frame based on the 1990 decennial census and was spread over 754 sample areas, which represent 2,007 geographic areas in the United States. Beginning in April 2004, the new sample drawn from Census 2000 was phased in over a 15-month period. From April 2004 to June 2005, the sample consisted of sample units drawn from both the 1990 and 2000 decennial censuses. Beginning in July 2005, the sample consists only of housing units drawn from Census 2000, along with housing units built after April 1, 2000. This sample is spread over 824 sample areas and represents the 2,025 geographic areas of the United States that existed at the time of Census 2000.

Beginning in the first quarter 1986, vacant seasonal mobile homes were included in the count of vacant seasonal units. This change resulted in a 12 percent increase in the number of vacant seasonal housing units.

Beginning with the first quarter 2002 Housing Vacancy Report, the size of the CPS/HVS sample increased to approximately 72,000 housing units. This expansion was one of the Census Bureau's plans to meet the requirements of the State Children's Health Insurance Program (SCHIP) legislation. Of the 72,000 housing units contained in the CPS/HVS sample, approximately 61,200 are eligible for interview each month; of this number, 3,900 occupied units, on the average, are visited but interviews are not obtained because occupants are not found at home after repeated calls or are unavailable for some other reason. In addition to the 61,200, there are also about 10,800 sample units in an average month which are visited but are found to be vacant or otherwise not to be interviewed. About half of the 10,800 are vacant and interviewed for the HVS.

The CPS estimation procedure for occupied units involves the inflation of the weighted sample results to independent estimates of the total civilian noninstitutional population of the United States by age, race, sex, and Hispanic/non-Hispanic categories. These independent estimates were based on statistics from the decennial censuses of population; statistics on births, deaths, immigration, and emigration; and statistics on the strength of the Armed Forces.

The HVS estimation procedure for vacant units is similar to that used for occupied units. Weighted sample results were adjusted at the state level using 2000 census vacant counts. A second adjustment inflated these results based on the CPS coverage of occupied units by geographic areas. As a final step for both the CPS and HVS, all housing unit counts were adjusted to reflect independent housing control totals. This change was effective, beginning in the first quarter 2003.

Data shown in all tables (except table 3) on vacancy rates and tenure of occupied units for the second quarter 2008 are from the CPS/HVS and are averaged for the three months, April, May, and June 2008. The data concerning the distribution of characteristics for occupied housing units, shown in table 3, were obtained primarily from the AHS national sample. Distributions of characteristics of occupied housing units from the AHS estimates were applied to CPS current housing inventory independent estimates to obtain the characteristics of occupied housing units used in this report. The Survey of Construction (SOC) and the Consumer Price Index also were used to improve estimates of the rent distribution.

The 2005 AHS sample was spread over 394 sample areas comprising 878 counties and independent cities with coverage in each of the 50 States and the District of Columbia. Of the 59,450 housing units both occupied and vacant contained in the AHS sample, 56,650 were interviewed and 6,150 were classified as "Type A noninterviews" for various reasons. 2,800 units were visited but were not eligible to be interviewed for the purposes of AHS. A detailed description of the AHS sample design and estimation procedure can be found in the H-150 report for 2005.

Comparability with Census of Housing Data

Most of the concepts and definitions were the same for items that appear in both the 1980 and 1990 censuses and the Housing Vacancy Survey. However, there was one minor difference in the housing unit definition between the CPS/HVS and the 1980 and 1990 decennial censuses. The difference was that, in the CPS/HVS prior to 1983, living arrangements containing five or more persons, not related to the person in charge, were classified as group quarters; for the 1980 and 1990 census, the requirement was raised to nine or more persons not related to the person in charge. For Census 2000, the conversion requirement was eliminated. There were some differences in what has been counted as housing units between the earlier censuses and the CPS/HVS. Descriptions of the differences between earlier censuses and the HVS appear in the 1985 and earlier reports of this series.

Prior to the first quarter 1990, there were significant differences between the CPS/HVS and the decennial censuses. The 1980 and 1990 decennial censuses included vacant mobile homes as housing units, whereas prior to 1986 the CPS/HVS did not. However, beginning in the first quarter 1986, vacant seasonal mobile homes were counted as housing units in the CPS/HVS. In addition, year-round vacant mobile homes were counted as housing units, beginning in the first quarter 1990 in the CPS/HVS. Another difference in the housing unit definition between the CPS/HVS (prior to 1986) and the 1980 and 1990 censuses was that the CPS/HVS required units to be separate living quarters and have direct access or have complete kitchen facilities. For the 1980 and 1990 decennial censuses, the complete kitchen facilities alternative was dropped with direct access required of all units. However, beginning in 1990, the CPS/HVS requirement for complete kitchen facilities was dropped with direct access required of all units. Thus, the earlier definitional differences were eliminated.

In addition, there are differences between the methodologies used to collect data for the CPS/HVS and the censuses. These differences include interviewing procedures, staff experience and training; differences in processing procedures and sample designs; the sampling variability associated with the CPS/HVS and the sample data from the census; and the non-sampling errors associated with the CPS/HVS and census data.

Research has shown that the CPS/HVS and the 1990 census produced significant differences for vacancy characteristics. The rental vacancy rate from the April 1990 census was 8.5 percent, whereas, the CPS/HVS reported the rental vacancy rate of 7.2 percent for the first half of 1990. The April 1990 census had a homeowner vacancy rate of 2.1 percent, while the CPS/HVS had a vacancy rate of approximately 1.7 percent for the first half of 1990. For occupied housing, the April 1990 census produced a homeownership rate of 64.2 percent, while for the first half of 1990 the CPS/HVS produced a rate of 63.9 percent. These differences illustrate that, for these characteristics as well as others, caution should be used when making comparisons between the 1990 census and the CPS/HVS.

Further research has shown that the CPS/HVS and the 2000 census produced significant differences for vacancy characteristics. The rental vacancy rate from the April 2000 census was 6.8 percent, whereas the CPS/HVS reported the rental vacancy rate of 7.9 percent for the first half of 2000. The April 2000 census had a homeowner vacancy rate of 1.7 percent, while the CPS/HVS had a vacancy rate of approximately 1.5 percent for the first half of 2000. For occupied housing, the April 2000 census produced a homeownership rate of 66.2 percent, while for the first half of 2000, the CPS/HVS produced a rate of 67.2 percent. These differences illustrate that, for these characteristics as well as others, caution should be used when making comparisons between the 2000 census and the CPS/HVS.

Comparability with Earlier Data

As stated earlier in this report, beginning with the first quarter 1994 new weighting procedures based on the 1990 decennial census were implemented. In addition, the survey data collection procedures became totally computerized. Caution should be used when comparing current data with unrevised data prior to 1994.

In the fourth quarter 1989, new edit procedures were implemented in the Current Population Survey/Housing Vacancy Survey (CPS/HVS). These new procedures were used to allocate cases that would have been classified as "not reported" under previous procedures.

In the first quarter 1990, year-round vacant mobile homes were included for the first time as part of the year-round vacant count of housing units. This change was made to make the composition of the housing unit inventory for the CPS/HVS similar to the decennial census and other surveys, which count all mobile homes as housing units when occupied or vacant (available for occupancy on the site). Research has shown that the inclusion of vacant mobile homes increased the vacancy rate significantly in some cases. We have revised 1989 data in this report to reflect all changes. Caution should be used when comparing data from 1990 or later with unrevised data prior to 1990.

In addition to the design and estimation changes mentioned in the previous section, caution should be used in comparing data for 1980 and beyond in this report with data from 1979 and earlier years. Starting with the first quarter 1980, several changes were implemented in the survey to improve the reliability of the data presented. These included adding a supplemental sample, refining the estimation procedures, and changing the source of occupied characteristics from the Quarterly Housing Survey to the AHS.

Although the above mentioned changes resulted in more reliable estimates, data for 1980 and later in this report are not completely comparable to data for the fourth quarter 1979 and previous quarters, as published in Housing Vacancies reports, series H-111, Nos. 1 to 79-Q4. Furthermore, unrevised data prior to 1990 were not completely comparable to 1990 data and beyond, due to the inclusion of year-round vacant mobile homes, beginning in the first quarter 1990. Thus, particular caution should be observed in drawing conclusions about trends that extend from before 1980 to 1980 and beyond, and also trends from before 1990 to 1990 and later. For comparative purposes, 1979 data in this report were revised to incorporate all changes made in 1980, and 1989 data were revised to incorporate all changes made in 1990. Unrevised 1989 and 1979 data are provided to show the magnitude of the various changes.

Vacancy Rates for Characteristics in Table 3

Vacancy rates in table 3 are based in part on forecasts of occupied housing units. These forecasts are periodically revised to incorporate more recent data and improved forecasting procedures. Beginning in the first quarter 2003, table 3 is based on data from the 2001 AHS. Beginning in 2005, first quarter 2004 data is based on data from the 2003 AHS. Beginnng in 2007, first quarter 2007 data is based on data from the 2005 AHS.

For the occupied unit forecasts for the monthly rent categories, we update the AHS data quarterly to reflect the rise in the cost of renting through the use of the residential rent index, and the latest available asking rent data for newly constructed rental units.

Caution in Using Seasonal Vacant Data

Analysis of seasonal vacant data prior to the first quarter 1987 has shown that estimates for these characteristics have been underestimated by approximately 28 percent. The estimates beginning with the first quarter 1987 were adjusted to reflect this.

Accuracy of the Estimates

Since the CPS/HVS estimates are based on a sample, they may differ somewhat from the figures that would have been obtained if a complete census had been taken using the same questionnaires, instructions, and enumerators. There are two types of errors possible in an estimate based on a sample survey: sampling and non-sampling. The accuracy of a survey result depends on both types of errors, but the full extent of the non-sampling error is unknown. Consequently, particular care should be exercised in the interpretation of figures based on a relatively small number of cases or on small differences between estimates. The standard errors provided for the CPS/HVS estimates primarily indicate the magnitude of the sampling error. They also partially measure the effect of some non-sampling errors in responses and enumeration; but do not measure any systematic biases in the data. (Bias is the difference averaged over all possible samples, between the estimate and the desired value.)

Non-Sampling Variability

Nonsampling errors can be attributed to many sources, e.g., inability to obtain information about all cases in the sample, definitional difficulties, differences in the interpretation of questions, inability or unwillingness on the part of respondents to provide correct information, inability to recall information, errors made in collection such as recording or coding the data, errors made in processing the data, errors made in estimating values for missing data, and failure to represent all units with the sample (undercoverage). Undercoverage in the CPS/HVS results from missed housing units and misclassifying housing units. Ratio estimation to independent controls, as described previously, partially corrects for the bias due to survey undercoverage. However, biases exist in the estimates to the extent that missed households have different characteristics than interviewed households.

Sampling Variability

The standard errors shown in the tables are primarily measures of sampling variability, that is, of the variations that occurred by chance because a sample rather than the entire population was surveyed. The sample estimate and its standard error enable one to construct confidence intervals; ranges that would include the average results of all possible samples with a known probability. For example, if all possible samples were selected, each of these being surveyed under essentially the same general conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then approximately 90-percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate would include the average result of all possible samples.

The average estimate derived from all possible samples is or is not contained in any particular computed interval. However, for a particular sample, one can say with specified confidence that the average estimate derived from all possible samples is included in the confidence interval.

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The most common types of hypotheses appearing in this report are: (1) the population parameters are identical, and, (2) the population parameters are different. An example of this would be comparing the vacancy rate in MAs versus the vacancy rate outside MAs. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

To perform the most common test, let x and y be sample estimates for two characteristics of interest. Let the standard error on the difference x-y be SEDIFF. If the ratio R = (x-y)/SEDIFF is between -1.645 and +1.645, no conclusion about the difference between the characteristics is justified at the 0.10 level of significance. If, on the other hand, this ratio is smaller than -1.645 or larger than +1.645, the observed difference is significant at the 0.10 level. In this event, it is a commonly accepted practice to say that the characteristics are different. Of course, sometimes this conclusion will be wrong. When the characteristics are, in fact, the same, there is a 10 percent chance of concluding that they are different. All statements of comparison in the text have passed a hypothesis test at the 0.10 level of significance or better. This means that, for most differences cited in the text, the absolute value of the estimated difference between characteristics is greater than or equal to 1.645 times the estimated standard error of the difference.

Comparisons of characteristics of the vacancies for the first quarter 1990, (which include year-round vacant mobile homes as part of the year-round vacant housing inventory for the first time) with previous unrevised quarters revealed significant differences in some cases. Thus caution should be used when comparing current data with unrevised data prior to 1990.

Illustrations of the Use of Tables of Standard Error

The sample estimate and its standard error enable one to construct a confidence interval. A confidence interval is a measure of an estimate's variability. The larger a confidence interval is in relation to the size of the estimate, the less reliable the estimate. For example, the estimated percent of all housing units vacant and available for rent is 3.1 percent (Table 10) and the standard error on the estimate is 0.1 percentage points (Table A-1). Then the 90-percent confidence interval is calculated as 3.1 percent + (1.645 x 0.1), or 3.1 percent + 0.2, or from 2.9 percent to 3.3 percent. If all possible samples were surveyed under essentially the same general conditions and the same sample design, and if an estimate and its standard error were calculated from each sample, then approximately 90 percent of the intervals from 1.645 standard errors below the estimate to1.645 standard errors above the estimate would include the average result of all possible samples.

Standard errors are also used to perform hypothesis testing--a procedure for distinguishing between population parameters (whose values are not known) using sample estimates. The most common type of hypothesis is that two population parameters are different. These tests may be performed at various significance levels. The significance level is the probability of concluding that the parameters are different when, in fact, they are the same. For example, to conclude that two parameters are different at the 0.10 level of significance, the absolute value of the difference between their corresponding sample estimates must be greater than or equal to1.645 times the estimated standard error of the difference.

The Census Bureau uses 90-percent confidence intervals and 0.10 levels of significance to determine statistical validity. The 90-percent confidence intervals are shown in the text or in the tables for selected items. The standard errors for other figures in this report are given in the tables. In addition to sampling error, the figures in this report, both the estimates and their standard errors, are also subject to rounding error.

Table A-1. Standard Errors of Percentages in Tables 10 and 11: Second Quarter 2009

Go to Housing Vacancies and Homeownership: Second Quarter 2009