American Housing Survey

# 2015 AHS Integrated National Sample:

# Sample Design, Weighting, and Error Estimation

Updated: September 6, 2018

---

U.S. Census Bureau, Department of Commerce
Department of Housing and Urban Development

# Contents

# 1. Overview

The purpose of this document is to describe the sample design, weighting, and error estimation for the 2015 American Housing Survey (AHS) Integrated National Sample.

For 2015, the U.S. Department of Housing and Urban Development (HUD) and the U.S. Census Bureau selected an entirely new sample for the AHS. The 2015 AHS sample is composed of an Integrated National sample and independent metropolitan area samples. The National sample is described as integrated because it incorporates a few different types of samples. The Integrated National sample includes—

- A representative sample of the nation.
- Representative oversamples of each of the 15 largest metropolitan areas.
- A representative oversample of HUD-assisted housing units.[1]

HUD and the Census Bureau intend to survey the entire Integrated National sample once every 2 years. As such, it is a longitudinal panel with a 2-year survey cycle.

The independent metropolitan area samples include representative samples of 10 selected metropolitan areas. For 2015, the 10 selected metropolitan areas represent one-half of what HUD and the Census Bureau refer to as the "Next 20" group of metropolitan areas (the second half will be included in the 2017 AHS). The Next 20 group of metropolitan areas is a subset of metropolitan areas ranging from the 16th to 50th largest, by population.[2] HUD and the Census Bureau intend to survey each member of the Next 20 group of metropolitan areas once every 4 years. As such, the Next 20 group of independent metropolitan area samples is a longitudinal panel with a 4-year survey cycle.

The Integrated National sample was interviewed between April 29 and September 11, 2015. The independent metropolitan area samples were interviewed at the same time, except for Phoenix, which was interviewed between July 27 and October 23, 2015.

# 2. Integrated National Sample Design

## 2.1. Eligible Universe

The universe of interest for the AHS consists of the residential housing units in the United States that exist at the time the survey is conducted. The universe includes both occupied and vacant units but excludes group quarters, businesses, hotels, and motels. Geographically, the survey covers 50 states and the District of Columbia.

## 2.2. Sample Size

For the 2015 AHS Integrated National sample, 85,393 sample housing units were originally selected for interview.[3] Of the selected units, 3,382 were found to be ineligible because the units either no longer existed or did not meet the AHS definition of a housing unit.

---

[1] Includes units in the public housing, project-based rental assistance, and Housing Choice Voucher programs.
[2] For more information about how the Next 20 group of metropolitan areas was selected, see "Metropolitan Area Selection Strategy: 2015 and Beyond" .
[3] In the 2014 document "Sample Sizes Determination and Decisions for the 2015 American Housing Survey and Beyond", Bucholtz and Ash discussed how the sample sizes were determined.

Of the 82,011 eligible sample units, about 12,500 (both occupied and vacant housing units) were classified as noninterviews because (1) no one was at home after repeated visits, (2) the respondent refused to be interviewed, or (3) the interviewer was unable to find the unit. This classification produced an unweighted overall response rate of 85 percent; the weighted overall response rate was also 85 percent.

Every sample unit of the 2015 Integrated National sample was asked a core set of questions. The sample also was randomly split into two samples, and each of these samples was asked a separate set of additional questions from four rotating topical modules. One set of the split samples was asked questions on the topical modules of housing counseling, arts and culture, and food security, while the other split sample set was asked questions on the topical module of healthy homes.

### 2.3. First Stage of Sample Selection: Select Primary Sampling Units

The Integrated National sample includes a representative sample of housing units for the United States. However, due to budget restrictions, the AHS does *not* include housing units from every political jurisdiction (that is, county or city) or census area (tract or block group) within the United States. To ensure the AHS sample is representative of all housing units in the United States, the Census Bureau adopted a sampling approach where they first selected *representative areas* throughout the United States (first-stage sampling), then they selected housing units within those areas (second-stage sampling).

The first stage of the sample selection was to determine which representative areas within the United States to include in the sample. To accomplish this, the United States was divided into areas made up of counties or groups of counties known as primary sampling units (PSUs),[4] of which there are two types: self-representing PSUs and non-self-representing PSUs.

A self-representing PSU is one that was included in sample with certainty. The sample from a PSU selected with certainty represents only that PSU, hence the term self-representing. The 2015 AHS sample included 85 self-representing PSUs—one for each of the 85 most populated Core Based Statistical Areas (CBSAs), where CBSA is defined according to the 2010 Office of Management and Budget as of February 2013.

A non-self-representing PSU is a PSU with a probability of selection of less than 1, meaning it may or may not be chosen to be in the sample. All counties outside of the top 85 largest CBSAs were grouped into 1,699 non-self-representing PSUs, which were further grouped into 224 strata. The Census Bureau then selected 1 non-self-representing PSU from each of the 224 strata to represent all PSUs in the strata. The probability of selection of a non-self-representing PSU from a stratum was proportional to the number of housing units in the PSU, relative to the number of housing units within the strata.

The end result of the first stage of sample section is an AHS sample spread over 309 PSUs. The 85 self-representing PSUs include 547 county and county equivalents, and the 224 non-self-representing PSUs consist of 353 county and county equivalents. Appendix A provides more details on the first stage of sample selection.

[4] The following are treated as county equivalents: independent cities in Maryland, Missouri, Nevada, and Virginia; boroughs in Alaska; and parishes in Louisiana.

## 2.4. Second Stage of Sample Selection: Select Housing Units Within PSUs

The second stage of sample selection involved selecting housing units from each of the 309 PSUs. The housing units were selected from a list of all housing units in the United States known as the Master Address File (MAF). The MAF is a data set maintained by the Census Bureau and based on updates from the prior decennial census and semiannual updates from the United States Postal Service (USPS) Delivery Sequence File, which itself consists of the addresses and mail routes serviced by the USPS. The MAF is updated semiannually in January and July, using information provided by the USPS. The 2015 AHS sample was based on the July 2014 MAF.[5]

The 2017 Integrated National Sample consisted of 85,393 housing units. To ensure the sample was representative of all housing units within the PSU, the Census Bureau stratified all housing units in each sample PSU into one of the following categories (known as strata):

- A HUD-assisted unit (as of 2013).
- Trailer or mobile home.
- Owner-occupied and one unit in structure.
- Owner-occupied and two or more units in structure.
- Renter-occupied and one unit in structure.
- Renter-occupied and two or more units in structure.
- Vacant and one unit in structure.
- Vacant and two or more units in structure.
- Other units, such as houseboats and recreational vehicles.

The information to create the stratification was based on the 2010 decennial census and a 2013 list of HUD-assisted units.[6] The sample rate for each stratum was constant, with the exception of the HUD-assisted strata, which was sampled at a higher rate to achieve the HUD requirement of approximately 5,250 HUD-assisted units in the sample.

---

[5] A small number of housing units (about 130) in remote rural areas, derived from another list known as the Coverage Improvement list, were added to the sample.

[6] In practice, the MAF was merged to both the 2010 decennial census and the 2013 HUD-assisted data, thereby permitting stratification of all housing units using the aforementioned housing characteristics.

## 3. Weighting

Each housing unit in the AHS sample represents itself and between 450 and 4,000 other units.[7] The exact number it represents is its "weight." The weight was calculated in four steps for two purposes: to minimize sampling errors and errors from incomplete data, and to force consistency with published estimates of certain housing and household characteristics that are believed to come from a more reliable data source.

### 3.1. Step 1: Base Weight Calculation

Every housing unit in the MAF had a positive probability of being selected into the AHS sample. The reciprocal of this probability of selection is referred to as the base weight and accounts for a sample housing unit's probability of selection in both the first and second stages of the National sample selection process.

### 3.2. Step 2: Total Housing Units Calibration for Non-Self-Representing PSUs

Each of the 224 non-self-representing PSUs represents all the PSUs within its respective stratum. The basic weight calculation in Step 1 reflects this feature of the sample design, but, in some cases, the AHS estimate of total housing units in each of the 224 strata and a Census Bureau independent estimate of housing units in the strata differ slightly. The Census Bureau adjusted for differences between these two estimates by calibrating the AHS basic weight from Step 1 to the independent estimate of housing units in the strata.

### 3.3. Step 3: Noninterview Adjustment Factor

Many eligible housing units selected for the AHS have potential respondents who do not complete an interview. Some are never home, refuse to answer, or had a language barrier, and sometimes, although rarely, the housing unit cannot be accessed by passable roads or the address cannot be found. These sample housing units result in a noninterview, which is also referred to as "unit nonresponse" and is different from "item nonresponse," which covers instances where an interviewee declines to answer a subset of AHS questions.

The noninterview adjustment factor (NAF) deals exclusively with unit nonresponse by expanding the weights of completed interviews to account for similar noninterviews. The calculation of the NAF involves three components—

1. Define NAF cells.
2. Calculate the NAF.
3. Collapse cells, if necessary.

Defining and calculating the NAF cells is a way of reducing the bias due to differential nonresponse. To reduce this nonresponse bias, the Census Bureau formed cells that include sample units that are homogenous to each other within the cells and heterogeneous between cells. Homogeneity and heterogeneity for sample units are measured with respect to the household's propensity to respond to the AHS interview.

[7] The mean value of the weights is 1,940. The median value of the weights is 1,309. The lower bound 5th percentile of weights is 464. The upper bound 95th percentile of weights is 3,982.

Table 3.1 summarizes the variables used in combination to define cells of the noninterview adjustment. Research conducted prior to 2015 determined the variables that best group sample units into cells with homogenous propensity to complete an AHS interview.

**Table 3.1: Variables Used to Define Noninterview Cells**

| Variable | Level Defined | Values |
|---|---|---|
| Census division | Division | Nine values |
| Type of housing unit | HU | (1) House, apartment or flat<br>(2) Mobile home<br>(3) Other |
| Core Based Statistical Area (CBSA; 2013) | County | (1) Metropolitan area: principal city<br>(2) Metropolitan area: nonprincipal city<br>(3) Micropolitan area<br>(4) Non-CBSA area |
| Quartiles of median income | Census block group | Four values for each of the four quartiles |
| Urban/rural status (2010)[a] | Area | (1) Urban (includes urban cluster or urban area)<br>(2) Rural |

[a] See https://www.census.gov/geo/reference/ua/urban-rural-2010.html for more information on the 2010 Urban and Rural classifications.

With the cells defined, the NAF within each cell is calculated as—

$$\text{NAF} = \frac{\text{Interviews} + \text{Noninterviews}}{\text{Interviews}}.$$

For both the numerator and the denominator of the NAF, weighted counts of the number of interviews and noninterviews were used. The counts were weighted using the product of the base weight (step 1) and housing unit calibration (step 2).

Lastly, cells of the NAF were collapsed if they have fewer than 25 sample housing units or the NAF is greater than 2.0. This avoided two potential problems: (1) unstable NAF estimates due to small cell counts and (2) large variances due to large adjustment factors.

It is important to note that some housing units selected for the AHS have respondents who complete enough questions in an interview for it to be considered a completed interview. However, if the respondents did not answer all the questions in the split sample modules, the housing unit is considered a noninterview for the split sample modules and will not have a value for the split sample weight.

### 3.4. Step 4: Housing and Demographic Adjustment Factors

The last step of calculating the weights is applying the Ratio Adjustment Factors (RAFs) to the weights to improve the coverage and reduce the variance of estimates. This step involves adjusting AHS weights to be consistent with known estimates of housing units and population from other data sources believed to be of superior quality or accuracy—these are referred to as "control totals." The RAF reduces the variance of an estimate when the control totals are associated with the estimated variable of interest.

The process of applying adjustment factor is called "raking." Ratio adjustments are a method of adjusting sample weights with control totals; their implementation is fairly straightforward.

$$\text{RAF} = \frac{\text{Independen t Estimate}}{\text{AHS sample estimate}}.$$

The calculation of the RAFs for AHS includes five steps:

1. Choose control totals and their adjustment priority order.
2. Define cells.
3. Calculate RAF iteratively, in order of importance (called raking).
4. Collapse cells.
5. Repeat raking until no further change is observed.

Table 3.2 provides information about the RAFs and their order of implementation. It is important to note the adjustment priority order reflects the importance placed on ensuring the AHS estimates, as adjusted, match the control totals. In other words, HUD and the Census Bureau place greater priority on adjusting AHS weights to match new construction control totals than the other control totals.

Additional information about the RAF and raking process is contained in Appendix B. Examples of the RAF process are contained in Appendix C.

**Table 3.2 Ratio Adjustment Factor Details**

| Adjustment Priority Order | Control Total | Cell Definition | Data Source(s) |
|---|---|---|---|
| 1 | Number of new construction housing units | Census Region and two categories of year built (2011–2012, 2013–2014) | Survey of Construction and Manufactured Homes Survey |
| 2 | Number of HUD-assisted housing units | Division/State/CBSA and three categories of HUD assistance (Public Housing, Private-Project Based, and Vouchers) | HUD |
| 3 | Number of total persons | Division/State/CBSA | Census Population Division |
| 4 | Number of Black persons | Division/State/CBSA | Census Population Division |
| 5 | Number of persons aged 65+ | Division/State/CBSA | Census Population Division |
| 6 | Number of Hispanic persons | Division/State/CBSA | Census Population Division |
| 7 | Number of total housing units | Division/State/CBSA | Census Population Division |

CBSA = Core Based Statistical Area. HUD = U.S. Department of Housing and Urban Development.

## 3.5. State-Level Weights – Within the AHS Integrated National Sample

Several states within the Integrated National Sample have large enough sample sizes to provide reasonable state-level estimates, including California, Florida, New York, Pennsylvania, and Texas. The criteria used to determine if a state can produce reasonable state-level estimates for a general set of state estimates were:

1. Key estimates have reasonable reliability. Reasonable reliability here means key estimates having Coefficients of Variance (CVs) less than 15% .
2. General estimates have reasonable reliability. For a representative set of estimates:
   - 50% or more of the estimates have CVs that are less than 15%, and
   - 10% or more of the estimates have CVs that are less than 5%.
   - The states of Georgia, Illinois, Massachusetts, and New Jersey were considered as well, but failed to meet these criteria.

No special state weight was derived for California, Florida, New York, Pennsylvania, and Texas, since the steps taken to improve the state estimates are part of the calculation of the weights for the Integrated National Sample.

## 3.6. State-Level Weights – Combination of the Integrated National Sample and the Independent Metropolitan Area Samples

In an effort to produce state-level estimates for additional states, sample from the Integrated National Sample and the Independent Metropolitan Area Samples was combined for select states. Using the criteria of section 3.5, Colorado and Ohio were able to support reasonable state-level estimates. With Colorado, we combined the Integrated National Sample of Colorado with the Independent Metropolitan Area Sample from the Denver CBSA. Likewise with Ohio, we combined the Integrated National Sample of Ohio with the Independent Metropolitan Area Sample from the Cleveland and Cincinnati CBSAs.

For sample housing units in Colorado and Ohio, a special state weight was calculated that is separate from the weights used with the Integrated National Sample and the Metropolitan Sample. This state weight combines the base weights of the two samples in order to minimize the overall variance of the resultant state estimates. With the combined base weights, noninterview adjustment factors and housing and demographic adjustment factors were applied that are analogous to the adjustment factors of the Integrated National Sample described in sections 3.3 and 3.4.

## 4. Nonsampling Errors

All numbers from the AHS, except for sample size, are estimates. As in other surveys, two types of general errors occur: sampling errors and nonsampling errors. Sampling errors are discussed in section 5. The definition of nonsampling errors is—

> Nonsampling errors arise mainly due to misleading definitions and concepts, inadequate sampling frames, unsatisfactory questionnaires, defective methods of data collection, tabulation, coding, incomplete coverage of sample units, and so on These errors are unpredictable and not easily controlled. Unlike sampling error, this error may increase with increases in sample size. If not properly controlled, nonsampling error can be more damaging than sampling error for large-scale household surveys.[8]

The various types of nonsampling errors are discussed in the following sections.

### 4.1. Coverage Errors

Coverage errors arise from the failure to give some units in the target population any chance of selection into the sample (undercoverage), or giving units more than one chance of selection (overcoverage). Because of deficiencies in the sampling lists, the housing units in the survey may not represent all housing units in the country. The Census Bureau attempts to address the deficiencies by adjusting the raw numbers from the survey proportionally so that the numbers published match independent estimates of the total number of housing units. This is part of the weighting production process described in section 3. Table 4.1 lists the sources of coverage errors. AHS users do not have to take any additional steps to account for coverage error.

[8] https://unstats.un.org/unsd/demographic/meetings/egm/Sampling_1203/docs/no_7.pdf

**Table 4.1 Sources of Coverage Errors**

| Type of Unit | Type of Coverage Error |
|---|---|
| Housing units with P.O. Box address or without 911 address[a] | The MAF includes 911 addresses because they can be located and does not include P.O. Box addresses. |
| New construction | Eligible units will be added but there is a lag between the time the unit is eligible and when it is added to the MAF. |
| Group quarters | Eligible units could be missed because of incorrect answers to questions used to screen out group quarters, which are ineligible units for the AHS. |

AHS = American Housing Survey. MAF = Master Address File.
[a] A number assigned to a structure that, in conjunction with a street or road name, identifies the location of the structure in the event of an emergency.

## 4.2. Nonresponse Error

Some respondents refuse the interview or cannot be located. HUD and the Census Bureau correct for nonresponse by implementing NAFs into the weighting process, as discussed in Section 3. AHS users do not have to take any additional steps to account for nonresponse error.

## 4.3. Measurement Errors from Missing Responses to Questions

Some respondents participate in an interview but refuse to answer questions or do not know a particular answer. For certain questions, HUD and the Census Bureau impute missing responses using various imputation techniques. The Census Bureau does not know how close the imputed values are to the actual values. For other items, "not reported" is used as an answer category. The items with the most missing data are primarily those that people forget or consider sensitive: mortgages, other housing costs, and income.

Incompleteness can cause large errors. A missing response in even 10 percent of sample units represents about 13.5 million homes (about 135 million homes are in the United States). To give users a sense of the bias caused by missing data, Table 4.2 provides estimates for Errors for Incomplete Data Bias.

**Table 4.2 Errors for Incomplete Data Bias for 2015 AHS-N (numbers in thousands)**

| When the AHS estimate of the number of housing units with a characteristic is… | …the chances are 90 percent that the actual value is within the range of plus or minus |
|---:|---:|
| 0 | 266 |
| 10 | 267 |
| 100 | 272 |
| 1,000 | 326 |
| 2,500 | 415 |
| 5,000 | 565 |
| 10,000 | 863 |
| 25,000 | 1,759 |
| 50,000 | 3,252 |
| 75,000 | 3,836 |
| 100,000 | 2,344 |
| 110,000 | 1,746 |
| 120,000 | 1,149 |
| 125,000 | 851 |
| 132,000 | 433 |
| 134,000 | 313 |
| 134,790 | 266 |

AHS-N = American Housing Survey – Integrated National Sample.

Table 4.2 is intended to be used only when a particular survey estimate is based on one or more variables with completeness rates of 50 to 90 percent. The values in Table 4.2 are based on a 1990 analysis by the Census Bureau, which estimated the standard error from missing data to be—

> .0012 x U + .0363 x (lesser of A or U-A), where A is any count of housing units with a characteristic from the AHS and U is the total number of housing units in the U.S. (134,790) or metropolitan area (both in thousands, result also in thousands).

Due to the large number of variables in the AHS, HUD and the Census Bureau typically do not publish completeness rates for individual survey estimates. AHS users who are interested in completeness rates should consider using the AHS public use file (PUF) microdata to estimate completeness rates. When using the PUF to estimate completeness rates, users should be aware of the following.

- PUF variables with a value of "not applicable" should not be considered missing. Not applicable means the question corresponding to the variable was not asked of the AHS respondent because the respond was not "in scope" for the question. For instance, if a respondent reported living in an apartment building, the respondent will not be asked questions about mobile home features.
- PUF variable with a value of "not reported" should be considered missing. A PUF variable will have a value of "not reported" if HUD and the Census Bureau did not develop an imputation process for the variable.
- For PUF variables for which HUD and the Census Bureau developed an imputation process, the variable will have a corresponding edit/imputation flag variable indicating whether the value of the variable was imputed for the respondent. The edit/imputation flag variables are the same as the variable name but are preceded by the letter "J." For instance, if a respondent

did not report a value for the variable HFUEL (heating fuel), but the respondent's value was imputed, the variable JHFUEL will equal "2," indicating an imputation.

- The edit/imputation flag will take a value of "1" if the respondent's reported value was edited. These edited values should not be considered missing.
- The correct way to calculate a completeness rate in the AHS is the following:

$$\frac{(\text{sum of respondents with reported values} - \text{sum of respondents with imputed values})}{(\text{sum of respondents with reported values} + \text{sum of respondents with "nonreported" values})}$$

### 4.4. Measurement Error from Inaccurate Responses to Questions

Wrong answers happen because people misunderstand questions, cannot recall the correct answer, or do not want to give the right answer. See *American Housing Survey for the United States: 2005*[9] for more discussion on this topic.

### 4.5. Question Validity Errors

In order to avoid the failure to design a survey question that accurately measures the construct of interest, HUD and the Census Bureau carefully test each new survey question to ensure it is measuring the construct of interest. Although some respondents possibly misinterpret a question, HUD and the Census Bureau do not have any additional information to estimate validity error rates. AHS users do not have to take any additional steps to account for validity error.

### 4.6. Processing Errors

After the data are collected, errors that can be introduced include data capture errors, data coding and classification errors, and data editing and imputation errors. HUD and the Census Bureau carefully test all aspects of the data capture, coding, classification, editing, and imputation procedures. Although mistakes are possible, HUD and the Census Bureau believe they are minimal. If a processing error is discovered, HUD and the Census Bureau will let AHS users know and, in some cases, will publish revised estimates. AHS users do not have to take any additional steps to account for processing error.

### 4.7. Additional Considerations

The AHS is a longitudinal survey conducted every 2 years. Many AHS users compare current-year AHS estimates with prior-year estimates. Users should be aware that HUD and the Census Bureau often make small changes to the text of various questions between surveys. AHS users comparing estimates with prior-year surveys should consult the document "Changes Between Surveys" that is published with each new AHS.

## 5. Sampling Errors

Error from sampling reflects how estimates from a sample vary from the actual value if all housing units had been interviewed under the same conditions. A confidence interval is a range that contains the actual value with a specified probability.

Users of the AHS PUFs can use replicate weights to create standard errors for any estimate. For further information, see "Guide to Estimating Variances Using Replicate Weights," which is

[9] https://www.census.gov/prod/2006pubs/h150-05.pdf

available on the Census AHS website (https://www.census.gov/programs-surveys/ahs/tech-documentation/help-guides/ahs-variance-estimation.html).

For users of the AHS Summary Tables, Generalized Variance Functions (GVFs) are a convenient tool for quick and easy estimation of sampling errors. The text below describes how to calculate sampling errors for counts, percentages, medians, and differences using GVFs.

## 5.1. Sampling Errors for Counts

Most published estimates from the AHS reflect weighted counts of housing units. The error from sampling for a weighted count is approximated using the following GVF for constructing a 90-percent confidence interval.

$$1.645 \times \sqrt{b \times A + a \times A^2},$$

where $A$ is the weighted count of housing units in thousands from the AHS and $a$ and $b$ are GVF parameters that vary depending on the characteristic being estimated.

Table 5.1 includes the values of $a$ and $b$ for six housing and household characteristics, plus a seventh set of values for all other characteristics. Users should note that values in Table 5.1 should be used to produce national estimates using the Integrated National sample, and are provided for use with full sample estimates and split sample estimates, and by occupancy type. Table 5.2 includes the values of $a$ and $b$ for types of vacant housing units, which should be used to produce National estimates using the Integrated National sample. Table 5.3 includes the values of $a$ and $b$ for use when producing estimates for the 15 metropolitan areas within the Integrated National sample.

**Table 5.1. General Variance Function Parameters for National Estimates**

| | Total and Total Occupied | | Owner Occupied | | Renter Occupied | |
|---|---|---|---|---|---|---|
| | ***a*** | ***b*** | ***a*** | ***b*** | ***a*** | ***b*** |
| **Full sample characteristic** | | | | | | |
| New construction | -0.000916 | 3.073 | -0.000709 | 3.240 | -0.001110 | 3.277 |
| Mobile homes | 0.000438 | 3.999 | 0.000233 | 3.442 | 0.000327 | 3.138 |
| Black | -0.000119 | 3.008 | -0.000055 | 3.231 | -0.000107 | 2.915 |
| Hispanic | -0.000095 | 2.430 | -0.000159 | 2.696 | -0.000099 | 2.383 |
| Elderly | -0.000085 | 3.067 | -0.000107 | 3.254 | -0.000031 | 2.406 |
| Below poverty | 0.000022 | 2.975 | 0.000042 | 3.090 | 0.000010 | 2.575 |
| All other | -0.000032 | 4.327 | -0.000042 | 4.271 | -0.000024 | 2.957 |
| | | | | | | |
| **Split sample characteristic** | | | | | | |
| New construction | -0.002014 | 6.404 | .002178 | 6.923 | -0.002260 | 6.468 |
| Mobile homes | 0.000443 | 6.630 | – .000317 | 6.167 | 0.000734 | 6.167 |
| Black | -0.000232 | 5.913 | .000121 | 6.125 | -0.000164 | 5.698 |
| Hispanic | -0.000187 | 5.052 | .000287 | 5.375 | -0.000211 | 5.196 |
| Elderly | -0.000159 | 5.613 | .000188 | 5.800 | -0.000096 | 5.188 |
| Below poverty | -0.000012 | 5.735 | – .000033 | 5.894 | -0.000071 | 5.371 |
| All other | -0.000052 | 7.060 | .000065 | 7.067 | -0.000049 | 5.983 |

**Table 5.2. General Variance Function Parameters for Seasonal and Vacant Housing Units for National Estimates**

| | Full Sample | | Split Sample | |
|---|---|---|---|---|
| **Type of Vacant Unit** | ***a*** | ***b*** | ***a*** | ***b*** |
| Seasonal | 0.001470 | 5.076 | 0.001334 | 7.764 |
| Total vacant | –0.000044 | 3.517 | –0.000071 | 6.058 |
| For rent | –0.000079 | 2.481 | –0.000062 | 4.774 |
| For sale only | –0.000092 | 2.576 | 0.000123 | 4.956 |
| Rented or sold | 0.000441 | 2.482 | 0.000679 | 4.906 |
| Occasional use/Usual residence elsewhere | 0.000340 | 3.730 | 0.000375 | 6.526 |
| Other vacant | –0.000123 | 2.897 | –0.000019 | 5.490 |

**Table 5.3 General Variance Function Parameters for Metropolitan Area Estimates**

| | | Full Sample | | Split Samples | |
|---|---|---|---|---|---|
| **Metropolitan Area (CBSA)** | **Domain** | *a* | *b* | *a* | *b* |
| Atlanta, GA | Total Units | -0.000364 | 0.81 | -0.000800 | 1.78 |
| | Owner Occupied | -0.000502 | 0.88 | -0.001005 | 1.88 |
| | Renter Occupied | -0.000502 | 0.90 | -0.001087 | 1.86 |
| Boston, MA | Total Units | -0.000439 | 0.83 | -0.001084 | 2.05 |
| | Owner Occupied | -0.000568 | 0.90 | -0.000918 | 2.01 |
| | Renter Occupied | -0.000773 | 0.92 | -0.000387 | 1.87 |
| Chicago, IL | Total Units | -0.000350 | 1.32 | -0.000851 | 3.21 |
| | Owner Occupied | -0.000503 | 1.44 | -0.001001 | 3.27 |
| | Renter Occupied | -0.000690 | 1.50 | -0.001544 | 3.40 |
| Dallas, TX | Total Units | -0.000307 | 0.82 | -0.000711 | 1.91 |
| | Owner Occupied | -0.000445 | 0.95 | -0.000899 | 2.02 |
| | Renter Occupied | -0.000443 | 0.93 | -0.000920 | 2.06 |
| Detroit, MI | Total Units | -0.000351 | 0.66 | -0.000772 | 1.46 |
| | Owner Occupied | -0.000449 | 0.70 | -0.000890 | 1.53 |
| | Renter Occupied | -0.000891 | 0.74 | -0.001720 | 1.56 |
| Houston, TX | Total Units | -0.000324 | 0.81 | -0.000707 | 1.77 |
| | Owner Occupied | -0.000519 | 0.95 | -0.000970 | 1.98 |
| | Renter Occupied | -0.000578 | 0.89 | -0.000982 | 1.88 |
| Los Angeles, CA | Total Units | -0.000345 | 1.57 | -0.001233 | 5.56 |
| | Owner Occupied | -0.000638 | 1.82 | 0.001355 | 3.80 |
| | Renter Occupied | -0.000508 | 1.61 | 0.000279 | 4.07 |
| Miami, FL | Total Units | -0.000325 | 0.81 | -0.001229 | 3.05 |
| | Owner Occupied | -0.000583 | 0.95 | 0.000063 | 2.21 |
| | Renter Occupied | -0.000481 | 0.92 | 0.000264 | 2.31 |
| New York, NY | Total Units | -0.000447 | 3.47 | -0.001141 | 8.85 |
| | Owner Occupied | -0.000663 | 3.64 | -0.000993 | 7.95 |
| | Renter Occupied | -0.000710 | 3.91 | -0.001029 | 8.50 |
| Philadelphia, PA | Total Units | -0.000353 | 0.86 | -0.000831 | 2.02 |
| | Owner Occupied | -0.000483 | 0.97 | -0.000954 | 2.11 |
| | Renter Occupied | -0.000691 | 1.00 | -0.000984 | 2.15 |
| Phoenix, AZ | Total Units | -0.000370 | 0.69 | -0.000786 | 1.46 |
| | Owner Occupied | -0.000522 | 0.76 | -0.001092 | 1.58 |
| | Renter Occupied | -0.000668 | 0.70 | -0.001101 | 1.46 |
| Riverside, CA | Total Units | -0.000335 | 0.52 | -0.000703 | 1.08 |
| | Owner Occupied | -0.000456 | 0.57 | -0.000957 | 1.18 |
| | Renter Occupied | -0.000482 | 0.52 | -0.001030 | 1.09 |
| San Francisco, CA | Total Units | -0.000346 | 0.62 | -0.000999 | 1.76 |
| | Owner Occupied | -0.000528 | 0.67 | -0.001166 | 1.63 |
| | Renter Occupied | -0.000596 | 0.68 | -0.001198 | 1.59 |
| Seattle, WA | Total Units | -0.000335 | 0.51 | -0.000783 | 1.20 |
| | Owner Occupied | -0.000535 | 0.59 | -0.000688 | 1.26 |
| | Renter Occupied | -0.000567 | 0.56 | -0.000571 | 1.27 |
| Washington, DC | Total Units | -0.000339 | 0.78 | -0.000837 | 1.93 |
| | Owner Occupied | -0.000488 | 0.86 | -0.000655 | 1.82 |
| | Renter Occupied | -0.000551 | 0.84 | -0.000555 | 1.85 |

As a hypothetical example with some assumed total, if a user wants to calculate the 90-percent confidence interval of the number of homes with some characteristic (assume *A* = 200), then the formula is—

$$1.645 \times \sqrt{4.327 \times 200 - 0.000032 \times 200^2} = 48.$$

The 90-percent confidence interval can then be formed by adding to and subtracting from this error to the survey estimate of 200 (that is, 200 plus or minus 48). There is a 90 percent chance that this interval contains the actual total and a 10 percent chance that it does not.[10]

Numbers in the published estimates are shown in thousands, so 200 means 200,000. The formulas are designed to use numbers directly from the published estimates; do not add zeros. The result is also in thousands, so 48 means 48,000.

### 5.2. Sampling Error for Percentages

Any subgroup can be shown as a percentage of a larger group. The error from sampling for a 90-percent confidence interval for this percentage is—

$$1.645 \times \sqrt{\frac{b \times p \times (100 - p)}{A}},$$

where *p* is the percentage; *A* is the weighted denominator, or base of the percentage in thousands; and *b* is the GVF parameter from Tables 5.1, 5.2 or 5.3.

For example, if a user wants to calculate the 90-percent confidence interval for a characteristic that represents 40 percent of the number of homes with a different characteristic (assume *A* = 200), the formula is—

$$1.645 \times \sqrt{\frac{4.327 \times 40 \times 60}{200}} = 11.9.$$

There is a 90 percent chance that the interval of 28.1 percent to 51.9 percent contains the true percentage.

Note that when a ratio *C*/*D* is computed where *C* is not a subgroup of *D* (for example, the number of Hispanic persons as a ratio of the number of Non-Hispanic Black persons), the error from sampling is different. The error from sampling for a 90-percent confidence interval for a ratio *C*/*D* is—

$$1.645 \times \left[\left(\frac{C}{D}\right)\sqrt{\left(\frac{\text{error for } C}{C}\right)^2 + \left(\frac{\text{error for } D}{D}\right)^2}\right].$$

[10] This formula gives 90-percent confidence interval errors. For 95-percent confidence interval errors, multiply by 1.96 instead of 1.645; for 99-percent confidence, multiply by 2.576 instead of 1.645.

The error for *C* should be interpreted as the error for a 90-percent confidence interval for *C*. Likewise, the error for *D* should be interpreted as the error for a 90-percent confidence interval for *D*.

### 5.3. Sampling Error for Differences

Two numbers from the AHS, like 21 and 34 or 34 percent and 55 percent, have a statistically significant difference if their 90-percent confidence intervals do not overlap.

When 90-percent confidence intervals do overlap, numbers are still statistically different if the result of subtracting one from the other is more than—

$$\sqrt{(\text{error for first number})^2 + (\text{error for second number})^2}.$$

The error for the first and second numbers should be interpreted as the error for a 90-percent confidence interval for the first and second numbers, respectively.

### 5.4. Sampling Error for Medians

Table 5.4 shows how to calculate the error from sampling for a 90-percent confidence interval for medians. This is an approximation of the error. The steps in Table 5.4 should only be used when the cumulative number of housing units for which the median applies is larger than 10 percent of the total number of housing units.

When cumulative number of housing units for which the median applies is smaller than 10 percent of the total number of housing units, the confidence interval on medians cannot be estimated reliably. To estimate a median's sampling error more accurately, use the steps in Table 5.5 to find the sampling error on 50 percent and apply it to compute the 90-percent confidence interval for the median.

The steps in Tables 5.4 and 5.5 are based on the hypothetical estimate reflected in Table 5.6.

**Table 5.4. Steps to Compute the 90-Percent Confidence Interval for a Median for Large Bases**

| Steps for Calculations | The Formula | A Hypothetical Example |
|---|---|---|
| How many total units is the median based on (in thousands, exclude "not reported" and "don't know")? | A | 200 |
| What is the estimated standard error of a 50-percent characteristic with a base equaling the total units? | $\sigma = \sqrt{\frac{4.5(0.5)(1-0.5)}{A}}$ | $\sqrt{\frac{4.5(0.5)(1-0.5)}{200}} = 0.075$ |
| What are the end points of the category the median is in? | X, Y | \$600, \$699 |
| What is the width of this category (in dollars, rooms, or whatever the item measures)? | W | \$100 |
| How many housing units are in this median category (in thousands)? | B | 30 |
| What is the estimated proportion of the total units falling in the category containing the sample median? | $P = \frac{B}{A}$ | $\frac{30}{200} = 0.15$ |
| Then the standard error from sampling for the median is approximately: | $se_{median} = \frac{\sigma \times W}{P}$ | $\frac{0.075 \times \$100}{0.15} = \$50$ |
| The 90-percent confidence interval for the median is: | $Median \pm 1.645 \times se_{median}$ | $Median \pm \$82.25$ |

**Table 5.5. Steps to Compute the Error From Sampling for a 90-Percent Confidence Interval for a Median for Small Bases**

| Item | Formula | Bottom Limit Hypothetical Example | Top Limit Hypothetical Example |
|---|---|---|---|
| How many total units is the median based on (in thousands, exclude "not reported")? | A | 200 | |
| Half the total, for the median (in thousands) | A/2 | 100 | |
| Error from sampling for 50 percent of the base of this median (first line) | $1.645 \times \sqrt{\frac{4.5(0.5)(1-0.5)}{\mathrm{A}}} = \frac{1.85}{\sqrt{\mathrm{A}}}$ | $\frac{1.85}{\sqrt{200}} = 0.131$ | |
| Multiply this percentage by total units to give the error in housing units. | $\frac{1.85}{\sqrt{\mathrm{A}}} \times \mathrm{A} = 1.85\sqrt{\mathrm{A}}$ | $0.131 \times 200 = 26.2$ | |
| Bottom of error range (second line minus fourth line, in thousands) | $\mathrm{B}_{\mathrm{bottom}}$ | 73.8* | |
| Top of error range (second line plus fourth line, in thousands) | $\mathrm{B}_{\mathrm{top}}$ | | 126.2* |
| *Start adding up the housing units in this table, category by category, cumulatively from the beginning of Table 5.6, until you exceed the starred number above. What interval does the starred number fall in? | | $500 – 599 | $700 – 799 |
| How many housing units are in all the categories before this one (in thousands)? | C | 50 | 125 |
| How many housing units are in this category (in thousands)? | D | 45 | 20 |
| What is the bottom limit of this category (in dollars, rooms, or whatever the item measures)? | E | $500 | $700 |
| What is the bottom limit of the next category (in dollars, rooms, etc.)? | F | $600 | $800 |
| Formula to calculate limits of confidence interval | $\frac{\mathrm{B}-\mathrm{C}}{\mathrm{D}}(\mathrm{F}-\mathrm{E}) + \mathrm{E}$ | $\frac{73.8-50}{45}(100) + 500$ | $\frac{126.2-125}{20}(100) + 700$ |
| Limits of confidence interval | | $552.89 | $706 |

*Starting with the starred step, this worksheet is equivalent to interpolation.

**Table 5.6. Hypothetical Household Income Median Estimate for Use with Tables 5.4 and 5.5 (numbers in thousands, except median)**

| | Number of Housing Units | Cumulative Number of Housing Units |
|---|---|---|
| **Total household monthly income** | 209 | |
| Less than $500 | 50 | 50 |
| $500 to $599 | 45 | 95 |
| $600 to $699 | 30 | 125 |
| $700 to $799 | 20 | 145 |
| $800 or more | 55 | 200 |
| Not reported | 9 | |
| | | |
| **Median Income**: $627 | | |

## 5.5. Additional Consideration

It should be noted that the minimum error from sampling is plus or minus 10 (meaning plus or minus 10,000).[11] If a formula gives an error smaller than 10, use 10.

[11] This minimum error formula is based on the following binomial 90-percent confidence interval on 0: $U \times (1 -.1^{4.33/U}) = 10$, (where $U$ is the total number of housing units from the AHS). For a 95-percent confidence interval, substitute .05 for .1 in the above formula. For a 99-percent confidence interval, substitute .01 for .1.

# Appendix A. Further Information on First Stage Sample Design

### Defining Primary Sampling Units

When counties are grouped into primary sampling units (PSUs), the three main requirements are:

1. PSUs never cross state boundaries.
2. The population of a PSU must be at least 7,500 people.
3. Generally, land area of PSUs should not exceed 3,000 square miles.

Requirements 2 and 3 reflect subtler difficulties encountered during the sample design. Requirement 2 addresses the problem that some counties lack enough housing units to ensure the American Housing Survey (AHS) will have enough sample units available throughout the lifetime of the survey, given the requirements of the Census Bureau.[12] Requirement 3 addresses the problem that some counties, particularly in the western United States and Alaska, are so geographically large and often contain difficult terrain with few main or secondary roads that travel within them poses significant practical burdens.

Requirement 2 took priority over requirement 3, which meant that every PSU had at least a population of 7,500, even if that meant having some PSUs with a land area greater than 3,000 square miles. Four PSUs in Alaska were the exceptions due to their large geographic size.

### Stratifying Non-Self-Representing PSUs

Stratification reduces the variance stemming from the first-stage sample design by grouping PSUs together that are homogeneous within the stratum and heterogeneous between strata with respect to variables of interest. This works because if units are alike within a given stratum, the overall estimate will not change much if different sample units are selected within the stratum.

To find the optimal set of strata the Census Bureau minimized the measure of homogeneity within strata, which was the weighted average of the coefficient of variation for eight totals associated with many AHS estimates. The eight variables chosen to measure homogeneity within first-stage strata are listed in Table A1.

---

[12] The design of the 2015 AHS sample was part of a much larger effort to redesign the demographic surveys of the Census Bureau that happens once a decade. The large demographic surveys included the Current Population Survey, Survey of Income and Program Participation, National Crime Victimization Survey, and the National Health Interview Survey. Statisticians from all surveys worked together to coordinate sample and share resources in the effort to redesign their surveys. Additionally, the surveys worked together to coordinate their respective samples with the goal that no housing unit in the United States will be selected for more than one of the aforementioned surveys within any 5-year time period.

**Table A1. Totals Chosen for Homogeneity Measures Within First-Stage Strata**

| **Statistic** |
|---|
| Urban housing units |
| Housing units with Hispanic head of household |
| Vacant-for-rent housing units |
| Owner-occupied housing units |
| Occupied housing units lacking complete kitchen facilities |
| Year-round housing units |
| Renter-occupied housing units |
| Households with low income<br>(householder income < 15th percentile in 2010 census) |

The following three constraints were used to create the non-self-representing (NSR) strata.

1. The largest NSR PSUs within a stratum could not be more than 10 percent larger in Measure of Size (MOS) than the smallest in the stratum.
2. Every stratum existed entirely within one of the nine census divisions.
3. Strata within census divisions were formed to be of equal size with respect to the MOS. This goal was accomplished by constraining the largest NSR stratum within a division to be no more than 20 percent larger than the smallest NSR stratum, wherein the MOS of a stratum is simply the sum of the MOS of every NSR PSU within the stratum.

While abiding to these constraints, an iterative algorithm was used to generate many possible sets of strata, and the set of strata that minimized the coefficient of variation within the eight variables shown in Table A1 was chosen as the final, defined set of strata for the national first-stage sample design.

# Appendix B. Ratio Adjustment Process Details

In the last step of calculating the weights, the Census Bureau applied the Ratio Adjustment Factor (RAF) to the weights to improve the coverage and reduce the variance of estimates. These goals were achieved by adjusting American Housing Survey (AHS) weights to be consistent with control totals of housing units and population. The RAF also reduces the variance of an estimate when the control totals are associated with the estimated variable of interest.

Generally speaking, ratio adjustments are a method of adjusting sample weights with control totals and their implementation is fairly straightforward. Take a control total, $X$, and its corresponding estimate, $\hat{X}$ , and multiply sample weights by a factor of $X / \hat{X}$ . This calculation results in adjusted sample weights that produce estimates that are much closer to the control total.

The calculation of the RAFs for AHS can be broken down into five steps.

1. Choose known totals.
2. Define cells.
3. Calculate RAF.
4. Collapse cells.
5. Repeat raking.

**Step 1. Choose Control Totals**

As mentioned previously, the Census Bureau wants control totals $X$ that are associated with the variable of interest. Control totals also require a reasonable corresponding estimate $\hat{X}$ from AHS. Both the control total $X$ and the AHS estimate $\hat{X}$ should define the same total. For example, a ratio adjustment for the total number of HUD housing units requires that both $X$ and $\hat{X}$ represent the same geographic area, apply to the same type of HUD program, and have the same reference period.

A second requirement for the control totals is that they should be a better estimate than the estimate produced from AHS. Again, these control totals are assumed to be more accurate than the AHS estimates and also have no variance.

Based on these two requirements, the following three data sources for control totals, described in Table B1, were considered to be suitable for ratios adjustments in 2015.

**Table B1. Sources of Control Totals**

| Control Total Candidates | Data Source |
|---|---|
| Number of housing units in HUD programs | HUD, based on 2015 HUD program data. |
| Number of new construction housing units | HUD and the Census Bureau, based on estimates from the 2011–2014 Survey of Construction and Manufactured Housing Survey, which were combined to define the new construction control totals. |
| Total population and housing unit counts by various characteristics | 2015 household population and housing unit demographic analysis projections derived from the 2010 census and estimated for July 1, 2015, by the Census Bureau Population Division. |

HUD = U.S. Department of Housing and Urban Development.

Given all possible control totals available in the sources listed in Table B1, HUD and the Census Bureau chose eight sets of totals within these three data sets to use for its RAF, as well as the priority order for which they are applied, which are presented in Table B2.

**Table B2. Ordered List of Control Totals**

| Order | Control Total and Source | Cell Definition |
|---|---|---|
| 1 | Number of new construction housing units | Region and Year built (2011–2012, 2013–2014) |
| 2 | Number housing units in HUD programs | Division/State/CBSA and HUD program (Public Housing, Private-Project Based, and Vouchers) |
| 3 | Number of total persons | Division/State/CBSA |
| 4 | Number of Black persons | Division/State/CBSA |
| 5 | Number of persons aged 65+ | Division/State/CBSA |
| 6 | Number of Hispanic persons | Division/State/CBSA |
| 7 | Number of total housing units | Division/State/CBSA |

HUD = U.S. Department of Housing and Urban Development.

## Step 2. Define Cells

Control totals within specifically defined groups of housing units, which are referred to as “cells,” were acquired for each of the chosen totals above. Estimates from the AHS were also calculated within these cells, and both of these were used to calculate RAFs.

Table B2 summarizes the cells for each set of ratio adjustments.

All of the ratio adjustments were applied at the division/state/county level except the new construction control totals, which were at the region/year level. Because Core Based Statistical Areas (CBSAs) can cross both division and state boundaries, the control totals were defined for the pieces of a CBSA in different divisions or states and for the remainder of the states and divisions.

*Cells defined by states.* As mentioned previously, several states had large enough sample sizes within the National sample to produce reasonable estimates. Ratio adjustments at the state level were included to improve the state-level estimates for California, Florida, New Jersey, New York, Pennsylvania and Texas. (Note that not all of these states ended up meeting the reasonableness criteria laid out in section 3.5)

*Cells defined by HUD programs.* The cells for the ratio adjustments of HUD housing units included three types of HUD programs: public housing, private-project based, and vouchers.

**Step 3. Calculate the RAF**

With the cells defined, the RAF for the first chosen control total—new construction—was calculated as—

$$\frac{\text{Control Total}}{\text{AHS Estimated Total}}.$$

This factor was then multiplied by the AHS weights to adjust AHS estimated counts within each cell. Ratio adjustments were applied iteratively using each of the remaining chosen control totals and their respectively defined cells in a process that is called raking. Each cell of each rake of Table B2 was adjusted with equation 1.

**Step 4. Collapsing RAF Cells**

RAF cells were collapsed for the same reasons noninterview adjustment factor cells were collapsed: (1) because a small number of sample housing units may produce an unstable estimate of the RAF and (2) to avoid large sample weights. To address both issues, cells are required to have at least 25 housing units, and the RAF must be less than or equal to 2.0. Cells were only collapsed after the first iteration of the raking through all of the chosen control totals in Table B2.

**Step 5. Repeat Raking**

After completing the first iteration of rakes and checking to see which cells need collapsing, raking was repeated using the ratios of chosen control totals over the modified AHS estimates until the AHS estimated totals stopped changing significantly between each raking step.

Appendix C provides a detailed example of how AHS uses raking within cells and across chosen totals.

# Appendix C. Examples of Ratio Adjustments

This appendix provides two hypothetical examples that demonstrate how the sample weights were adjusted so that they were consistent with a set of control totals. The first example is a ratio adjustment, and it is provided as context because it is a special case of raking—one rake. The second example demonstrates how to complete a more complicated raking adjustment.

For the two examples, assume weights were calculated for a sample and the weights included all weighting adjustments up to a nonresponse adjustment. With these weights, hypothetical totals by two categories of tenure status (owner or renter) and two categories of type of construction (old or new) are assumed. Table C1 summarizes the estimated totals resulting from this hypothetical sample and weights.

**Table C1: Estimated Totals**

| | Owners | Renters | Total |
|---|---|---|---|
| New | 110 | 91 | 201 |
| Old | 97 | 107 | 204 |
| Total | 207 | 198 | 405 |

### Example 1: Ratio Adjustment

Suppose the control totals were as shown in Table C2.

**Table C2: Example 1 Control Totals**

| | Owners | Renters | Total |
|---|---|---|---|
| New | 115 | 105 | 220 |
| Old | 95 | 105 | 200 |
| Total | 210 | 210 | 420 |

The control totals of Table C2 are used to improve the weights by making the estimates from the weights consistent with the control totals. Table C3 shows the Ratio Adjustment Factor (RAF) that will make the estimated totals consistent with the control totals.

**Table C3: Example 1 Ratio Adjustment Factors**

| | Owners | Renters |
|---|---|---|
| New | 115/110 = 1.0455 | 105/91 = 1.1583 |
| Old | 95/97 = 0.9794 | 105/107 = 0.9813 |

If the factors from Table C3 are applied to the weights of the sample units, then the estimates from the revised weights will be consistent with the totals of Table C2.

Note that ratio-adjusted weights for the combination of owners and new construction is the product of the weight before raking with the RAF, that is,

$$\text{Ratio-adjusted weight} = \text{original weight} \;\square 1.0455 .$$

The ratio-adjusted weights for the other three cells are defined similarly.

**Example 2: Raking Adjustment**

Table C4 shows different control totals than those of Table C2.

**Table C4: Example 2 Control Totals**

| | Owners | Renters | Total |
|---|---|---|---|
| New | ? | ? | 220 |
| Old | ? | ? | 200 |
| Total | 210 | 210 | 420 |

Table C4 does not have the totals for the specific combinations of tenure status and old or new construction; however, totals can be used with raking to improve the weights.

Raking is the repeated application of ratio adjustments to the marginal totals. Ratio adjustments are repeated for each set of marginal totals—the row totals and the column totals in this example. It can be shown that raking will converge to a unique solution.

First, raking the categories of old or new construction is done. This involves adjusting the cells for the totals of old or new construction. Table C5 shows the calculated adjustment factors for the first rake.

**Table C5: Factors for First Rake—Old or New Construction**

| | Ratio | Factor |
|---|---|---|
| New | 220/201 | 1.0945 |
| Old | 200/204 | 0.9804 |

For new construction, the value of 220 came from the marginal control total of new construction (first row) in Table C4, and the value of 201 came from the marginal estimated total of new construction (first row) in Table C1.

The ratios of table C5 are then applied to the totals, or, equivalently, the weights of the sample units that are used to calculate the total. Table C6 shows the application of the factors from Table C5 to the totals of Table C1.

**Table C6: New Total for First Rake—Old or New Construction**

| | Owners | Renters |
|---|---|---|
| New | 110 x 1.0945 = 120.40 | 91 x 1.0945 = 99.60 |
| Old | 97 x 0.9804 = 95.10 | 107 x 0.9804 = 104.90 |

Table C7 shows the result of the first rake—the application of the factors from Table C6 to the totals of Table C1.

**Table C7: Revised Totals for First Rake—Old or New Construction**

| | Owners | Renters | Total |
|---|---|---|---|
| New | 120.40 | 99.60 | 220.00 |
| Old | 95.10 | 104.90 | 200.00 |
| Total | 215.50 | 204.50 | 420.00 |

After the first rake, the revised estimates are now consistent with the old or new construction column totals, but the estimated row totals are not consistent with the tenure control totals.

The tenure totals are then raked using the revised totals in Table C7. The ratio adjustments are calculated with the revised tenure totals from Table C7 and the control totals from Table C4. Table C8 shows the factors needed to adjust the owner or renter columns.

**Table C8: Factors for Second Rake—Tenure**

| | Ratio | Factor |
|---|---|---|
| New | 210/215.51 | 0.9745 |
| Old | 210/204.505 | 1.0269 |

The ratios of Table C8 are then applied to the weights of the sample units within owners and renters in Table C9.

**Table C9: New Totals for Second Rake—Tenure**

| | Owners | Renters |
|---|---|---|
| New | 120.45 x 0.9745 = 117.33 | 99.645 x 1.0269 = 102.28 |
| Old | 95.06 x 0.9745 = 92.67 | 104.86 x 1.0269 = 107.75 |

Table C10 shows the complete result of the second rake—the application of the factors from Table C9 to the totals of Table C7.

**Table C10: Revised Totals for Second Rake—Tenure**

| | Owners | Renters | Total |
|---|---|---|---|
| New | 117.33 | 102.28 | 219.61 |
| Old | 92.67 | 107.72 | 200.39 |
| Total | 210.00 | 210.00 | 420.00 |

With the second rake, the revised estimates are now consistent with the tenure row totals, but the estimated row totals are not consistent with the tenure control totals. However, both the row and the column totals are closer to the control totals. A third rake is done to adjust for the old or new construction totals again. Table C11 shows the factors of the third rake, and Table C12 shows the resultant totals.

**Table C11: Factors for Third Rake—Old or New Construction**

| | Ratio | Factor |
|---|---|---|
| New | 220/219.61 | 1.0018 |
| Old | 200/200.39 | 0.9980 |

**Table C12: Revised Totals for Third Rake—Old or New Construction**

| | Owners | Renters | Total |
|---|---|---|---|
| New | 117.54 | 102.46 | 220.00 |
| Old | 92.49 | 107.51 | 200.00 |
| Total | 210.03 | 209.97 | 420.00 |

The fourth rake repeats the adjustment for the tenure totals. Table C13 shows the factors of the third rake, and Table C14 shows the resultant totals.

**Table C13: Factors for Fourth Rake—Tenure**

| | Ratio | Factor |
|---|---|---|
| Owner | 220/210.03 | 0.9999 |
| Renter | 200/2009.97 | 1.0001 |

**Table C14: Revised Totals for Fourth Rake—Tenure**

| | Owners | Renters | Total |
|---|---|---|---|
| New | 117.52 | 102.48 | 220.00 |
| Old | 92.48 | 107.52 | 200.00 |
| Total | 210.00 | 210.00 | 420.00 |

Table C14 shows the final result of the raking. The original estimated totals are now revised so that both the row totals and column totals are consistent with the control totals of old or new construction and tenure.

To clarify how this applies to the weights, note that raking-adjusted weights for the combination of owners and new construction is the product of the weight before raking with the factors of the four rakes, that is,

$$\begin{aligned}\text{Raking-adjusted weight} &= \text{original weight} \times 1.0945 \times 0.9745 \times 1.0018 \times 0.9999 \\ &= \text{original weight} \times 1.0684 .\end{aligned}$$

The raking-adjusted weights for the other three cells of example 2 were done similarly.

The adjustment factors in the tables were displayed with rounding to four decimal points. No rounding is done in the actual calculation of the raking prior to their application because the raking would not converge if the factors were rounded.