

Survey of Market Absorption: Source and Accuracy Statement for 2018 Completions

Overview

The purpose of this document is to describe the sample design, weighting, and accuracy of the estimates for the Survey of Market Absorption (SOMA) for buildings constructed in 2018.

The SOMA is designed to provide estimates of the rate at which privately financed, nonsubsidized units in buildings with five or more units are rented or sold (absorbed).

Eligible Universe

The universe of interest for the SOMA consists of the units within privately-financed multifamily residential buildings with five or more units. SOMA does not include buildings completed in areas that do not issue permits. Less than two percent of all new construction takes place in non-permit areas¹.

Sample Design

Buildings for SOMA come from those included in the U.S. Census Bureau's Survey of Construction (SOC). Since SOMA uses the sample from SOC, we begin by describing the sample design of SOC.

For the first-stage sample design of SOC, the United States was first divided into primary sampling units (PSUs) that were stratified based on population and building permit activity. If a PSU had a large population or high permit activity, it was classified as self-representing. Otherwise it was classified as non-self-representing. There were 48 self-representing PSUs and 772 non-self-representing PSUs. The non-self-representing PSUs were stratified into 121 strata by Census Division, permit activity, metropolitan status, and population. One PSU was then randomly selected from each stratum.

In the second-stage sample design, a sample of geographic locations was chosen within each of the 169 selected PSUs. Areas that do not require building permits were selected separately from the permit-issuing places. Approximately 80 block groups of non-permit areas were selected within the sample PSUs. The permit-issuing places were stratified by permit activity. Approximately 900 permit-issuing places were selected.

In the third stage of the sample design, permits are selected monthly. Within sampled places, permits for buildings with five or more units are selected with certainty and permits for buildings with one to four units are subsampled. In the non-permit areas, field representatives canvass the areas and all housing unit construction starts are included in the sample with

¹ Building Permits Survey coverage, <https://www.census.gov/econ/overview/co0200.html>

certainty. For further details on the SOC sample design and weighting see https://www.census.gov/construction/nrc/how_the_data_are_collected/soc.html.

When SOC indicates the construction of the building is complete², that building is selected for SOMA in order to collect absorption data. All buildings with five or more units are selected for SOMA, except when the monthly sample size is larger than 1,200, in which subsampling is used to reduce the sample (Alexander 1998). Historically, subsampling has not been applied to SOMA since 2005.

In intervals of three, six, nine, and twelve months after completion, the SOMA respondent reports how many units were rented or sold (i.e., absorbed) in each time period.

Types of Estimates

The SOMA produces three types of estimates: quarterly estimates, estimates for the prior four quarters, and annual estimates.

1. Quarterly Estimates

Each quarter, the SOMA publishes estimates of the number of completed units and the proportion of those units absorbed in the latest quarter, and revises estimates published in the three preceding quarters.

- a. For units completed in the latest quarter, the three-month absorptions are estimated.
- b. For units completed in the first-previous quarter, the six-month absorptions are estimated, along with revisions to the previously-published estimates.
- c. For units completed in the second-previous quarter, the nine-month absorptions are estimated, along with revisions to the previously-published estimates.
- d. Lastly, for units completed in the third-previous quarter, the twelve-month absorptions are estimated, along with final revisions to the previously-published estimates.

After the fourth quarter, the estimates are no longer revised. In the quarterly tables, estimates are made for totals, absorption rates, and median unit price.

2. Estimates for the Prior Four Quarters

For the reported quarter, the SOMA publishes estimates on the units absorbed in that quarter for the completed buildings of the prior four quarters including:

² Construction of a multifamily building is classified as complete when 50 percent of the units are available for occupancy. New Residential Construction, <https://www.census.gov/construction/nrc/definitions/index.html>

- a. number of units completed in the four quarters prior to the reported quarter
- b. number of units absorbed prior to the current quarter
- c. number of units absorbed in the current quarter
- d. number of units not absorbed (still available for sale or rent) at the end of the current quarter

The same sample units used in the quarterly estimates are used to produce the estimates for the prior four quarters. However, the prior four quarters' estimates aggregate the survey results of four quarters to focus on the absorption activity in the current quarter, while the quarterly estimates focus on cumulative absorption activity within each of the four quarters. In the estimates for the prior four quarters, estimates are made for totals and median unit price.

3. Annual Estimates

Once a year, the SOMA produces calendar-year estimates of the number of units completed and the proportion of those units absorbed after three, six, nine, and twelve months. All units in the estimates have been interviewed through twelve months, and therefore the final estimates are used from each of the four quarters. In the annual tables, estimates are made for totals, absorption rates, and median unit price.

Weighting

Each building in the SOMA sample represents itself and other buildings that were not selected from the given stratum. The exact number it represents is its "weight." The weight is calculated in two steps. The purpose of these steps is to minimize both sampling and nonsampling error.

1. **Basic weight.** The Census Bureau assigned each unit a weight to reflect its probability of selection. Any subsampling adjustments are included in the basic weight.
2. **Raking to Known Totals.** The basic weights are raked to two sets of SOC known totals: Regional and Metropolitan Area. To ensure convergence, the ratio adjustments are iterated three times.
 - a. **Survey of Construction regional ratio adjustment.** The Census Bureau aligns the quarterly SOMA sample estimate of the number of units in privately-financed residential buildings with five or more units for each census region with the quarterly SOC sample estimate of the number of units in privately-financed residential buildings with five or more units for each census region by multiplying the basic weight by the following ratio adjustment factor:

$$\frac{\text{SOC estimate of the number of units in privately-financed residential buildings with five or more units for quarter } q, \text{ region } r}{\text{SOMA estimate of the number of units in privately-financed residential buildings with five or more units for quarter } q, \text{ region } r}$$

Although the SOMA sample is a subset of the SOC sample, the two estimates may differ because (1) the SOC also calculates weighting adjustments that cannot be captured in the SOMA basic weight, (2) there are sample buildings that enter the SOMA sample too late to be included in the preliminary SOMA estimates, (3) some buildings are found to be ineligible for the SOMA after construction is complete, and (4) the SOC sample estimate may contain buildings completed in non-permit areas, though construction of buildings with five or more units in non-permit areas is rare.

- b. Survey of Construction metropolitan status adjustment.** The Census Bureau aligns the quarterly SOMA sample estimate of the number of units in privately-financed residential buildings with five or more units by inside/outside metropolitan area with the quarterly SOC sample estimate of the number of units in privately-financed residential buildings with five or more units by inside/outside metropolitan area by multiplying the basic weight by the following factor:

$$\frac{\text{SOC estimate of number of units in privately-financed residential buildings with five or more units for quarter } q, \text{ metropolitan status } m}{\text{SOMA estimate of number of units in privately-financed residential buildings with five or more units for quarter } q, \text{ metropolitan status } m}$$

Accuracy of Estimates

The SOMA is a sample survey and consequently all estimates in this report are subject to sampling variability. Since the statistics produced from the SOMA are estimates derived from a sample, they will differ from the “true values” being estimated. There are two types of errors, which cause estimates based on a sample survey to differ from the true value: sampling error and nonsampling error. We next discuss these two types of error.

Nonsampling Errors

If every building in the United States were interviewed, the estimates of housing unit characteristics could still differ from the true value due to nonsampling errors. In general, nonsampling errors can be attributed to many sources: inability to obtain information about all cases in the sample, difficulties with definitions, differences in interpretation of questions, inability or unwillingness of the respondents to provide correct information, and errors made in processing the data.

For the SOMA we produce response rates as a measure of nonsampling error. The response rates for the SOMA are calculated by dividing the weighted number of building interviews by the weighted number of eligible buildings:

$$\text{Response Rate} = \frac{\text{Buildings with a completed Interview}}{\text{Eligible buildings}} \times 100$$

Response rates are calculated for the initial interview and are reviewed for data quality issues. They are also revised and published after the fourth and final interview.

An eligible building means that construction of the building was (1) privately-financed, (2) the construction was for residential purposes, and (3) the building contains five or more units. Only the responses from completed interviews are used to produce the SOMA estimates.

A survey interview is considered complete once either of these conditions is met:

- The field representative collects information about the number of bedrooms and their cost information for all nonsubsidized apartment units or condominium/cooperative units in the building.
- The units in the building are timeshares, public housing units³, subsidized units (e.g., Section 8 or any other government assistance), townhouse units, or cash-and-carry units (i.e., rent is paid up-front). The SOMA universe consists of privately-financed residential buildings with five or more units. Although these types units are constructed with private financing, they are not included in SOMA absorption estimates.

Although no other direct measurements of nonsampling error have been obtained, the Census Bureau has determined that most response and operational errors are detected and corrected during review of the data for reasonableness and consistency.

Table 1 provides the final response rates for each completion quarter from Q1 through Q4 2018, with an overall annual rate for 2018.

Table 1. Survey of Market Absorption Quarterly Response Rates

Completion Period	Sample Size	Eligible		Ineligible	Weighted Response Rate
		Interviews	Noninterviews		
2018 annual	4,000	3,900	70	30	98.1
Q4 2018	900	900	20	N<15	97.6
Q3 2018	1,000	1,000	N<15	N<15	98.5
Q2 2018	1,100	1,100	20	N<15	97.6
Q1 2018	950	950	N<15	N<15	98.6

Source: U.S. Census Bureau, 2018-2019 Survey of Market Absorption.

Historical response rates for 2014 to 2018 are provided in Appendix A.

³ Public housing buildings constructed with any private funds are part of the SOMA universe, while those constructed solely with public financing are part of the Value of Construction Put in Place (VIP) survey universe.

Sampling Errors

The particular sample used for this survey is one of many possible samples that could have been selected using the same design. Even if the same questionnaires, instructions, and interviewers were used, estimates from each of the different samples would likely differ from each other. The deviation of a sample estimate from the average from all possible samples is defined as the sampling error. The standard error of a survey estimate provides a measure of this variation and, thus, is a measure of the precision with which an estimate from a sample approximates the average result from all possible samples.

As calculated for this survey, the standard error also partially measures the variation in the estimates due to errors in responses and by the interviewers (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on the standard error, biases, and some additional nonsampling errors not measured by the standard error. As a result, the standard error of estimates based on this sample reflect only a portion of the uncertainty that actually exists. Nonetheless, they are extremely useful because they capture all of the effect of sampling error and, in this case, some nonsampling error as well.

If all possible samples were selected, if each of them was surveyed under the same general conditions, if there were no systematic biases, and if an estimate and its estimated standard error were calculated from each sample, then:

- Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate (i.e., the 68 percent confidence interval) would include the “true value.”
- Approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate (i.e., the 90 percent confidence interval) would include the “true value.”
- Approximately 95 percent of the intervals from 1.96 standard errors below the estimate to 1.96 standard errors above the estimate (i.e., the 95 percent confidence interval) would include the “true value.”

The 90 percent confidence level is standard for measuring statistical significance. A Margin of Error is a measure of an estimate’s reliability. The Margin of Error (MOE) is the distance between an estimate and its lower or upper confidence bound. The larger the MOE in relation to the size of the estimate, the less reliable the estimate. This number, when added to and subtracted from the estimate, forms the 90 percent confidence interval.

Two estimates have a statistically significant difference if their margins of error (MOEs) do not overlap. When MOEs do overlap, estimates are still significantly different if the result of subtracting one from the other is more than

$$\sqrt{(\text{MOE for first estimate})^2 + (\text{MOE for second estimate})^2}$$

For very small estimates, the lower limit of the confidence interval may be negative. In this case, a better approximation to the true interval estimate can be achieved by restricting the interval estimate to positive values; that is, by changing the lower limit of the interval estimate to zero.

The reliability of an estimated absorption rate (i.e., a percentage) calculated by using sample estimates for both the numerator and denominator depends on both the size of the rate and the size of the total on which the rate is based. Estimated rates of this kind are more reliable, relative to the size of the rate, as both components increase in size. Estimated rates of this kind are relatively more reliable than the corresponding estimates of the numerators of the rates, particularly if the rates are 50 percent or more.

Sampling Error Estimation

The SOMA uses replication methods to produce direct estimates of sampling error, although dissemination of these variance estimates has changed. Starting with Second Quarter 2018 completions, the SOMA is publishing MOEs for every estimate given in the Table Creator⁴ application. Prior to the publication of the estimates for the Second Quarter 2018 completions, the Department of Housing and Urban Development (HUD) and the Census Bureau published estimates of sampling errors in two ways depending on the type of publication.

1. For reports published quarterly, HUD and the Census Bureau provided data users with tables of estimates and corresponding MOEs for each estimate (U.S. Census Bureau 2018a). The variance estimates were produced with replication methods.
2. For reports published annually, HUD and the Census Bureau provided data users with tables of estimates and abbreviated standard error tables, along with instructions to interpolate a standard error estimate and calculate the resulting MOE (U.S. Census Bureau 2018b). The variance estimates produced for these abbreviated tables were calculated using a Generalized Variance Function (GVF). The GVF is a model that produces a predicted standard error for a given sample estimate.

References

Alexander, C. (1998). Census Bureau Memorandum, "Survey of Market Absorption (SOMA): Specifications for Selecting the Monthly Sample," from Charles H. Alexander, Jr. to Barbara M. Walter, working draft dated October 1, 1998.

U.S. Census Bureau (2017). *Survey of Market Absorption of New Multifamily Units, 2017 Annual Absorption Report (2016 Absorptions for 2015 Completions)*. April 2017.

⁴ Hyperlink to the Table Creator application: https://www.census.gov/data-tools/demo/soma/soma.html?s_appName=soma&s_tableName=TABLE1&s_byGroup1=0&s_unit=R

U.S. Census Bureau (2018a). *Survey of Market Absorption of New Multifamily Units, Second Quarter 2018 - ABSORPTIONS (Completions in First Quarter 2018)*. September 2018.

U.S. Census Bureau (2018b). *Survey of Market Absorption of New Multifamily Units, 2018 Annual Absorption Report (2017 Absorptions for 2016 Completions)*. April 2018.

Miscellaneous

File name: 2018 SOMA-Source and Accuracy-v1.0.docx

Table A1. Historical Survey of Market Absorption Response Rates

Completion Period	Sample Size	Eligible		Ineligible	Weighted Response Rate
		Interviews	Noninterviews		
2018 annual	4,000	3,900	70	30	98.1
Q4 2018	900	900	20	N<15	97.6
Q3 2018	1,000	1,000	N<15	N<15	98.5
Q2 2018	1,100	1,100	20	N<15	97.6
Q1 2018	950	950	N<15	N<15	98.6
2017 annual	4,400	4,200	100	50	97.4
Q4 2017	1,100	1,000	30	N<15	96.5
Q3 2017	1,200	1,100	40	30	96.7
Q2 2017	1,100	1,100	40	N<15	97.9
Q1 2017	950	950	20	N<15	98.7
2016 annual	4,300	4,000	60	150	98.4
Q4 2016	1,200	1,100	N<15	40	99.3
Q3 2016	1,200	1,100	20	50	98.6
Q2 2016	1,000	950	N<15	50	97.8
Q1 2016	950	900	20	20	97.8
2015 annual ⁵	4,100	3,900	100	60	97.9
Q4 2015	1,000	1,000	30	20	97.0
Q3 2015	1,200	1,200	40	20	97.5
Q2 2015	1,000	950	20	N<15	98.6
Q1 2015	800	800	20	N<15	98.6
2014 annual	3,600	3,500	30	60	99.4
Q4 2014	900	850	N<15	N<15	99.0
Q3 2014	1,000	1,000	N<15	20	99.3
Q2 2014	900	850	N<15	N<15	99.6
Q1 2014	800	800	N<15	N<15	99.6

Source: U.S. Census Bureau, 2014-2018 Survey of Market Absorption.

⁵ The response rates for 2015 differ from what was published in the 2017 Annual Report for 2015 completions (U.S. Census Bureau 2017) due to changes in definitions of eligible units and respondents in the response rate calculation. Response rates had not been published in the Annual Report prior to 2017.