

Updating the Renter Sample for the CPI Housing Survey

By Frank Ptacek

The CPI Housing Survey provides the data needed to measure price change for the two CPI Housing components: *Owners' equivalent rent of primary residence* (OER) and *Rent of primary residence* (RENT). These are the largest CPI components with 22.1 and 6.4 percent of the CPI weight, respectively. The Housing Survey follows the rents of a sample of renter-occupied housing units selected to represent both renter and owner-occupied housing units in the urban United States. The CPI has 87 geographic Pricing Areas selected to represent the urban United States, and BLS field staff based in those areas collect data for the units of the Housing Survey sample on a regular, recurring basis.

Until 2002, the CPI Shelter Index was revised as part of a periodic roughly 10-year cycle during which all dimensions of the CPI (expenditure patterns, items and stores, rented housing units, urban areas for collection and computation and collection methods) were revised. The last comprehensive CPI revision took place in 1998. Beginning in 2002, the BLS replaced these comprehensive periodic revisions and their associated large resource spikes with shorter, more continuous updating. For example, expenditure patterns are now updated every two years, items and outlets priced are updated over a four-year period, and computer systems and collection methods are regularly maintained. The only dimensions not included in the new revision paradigm were the rented housing units used to measure shelter and the urban areas in which data is collected for the monthly CPI.

The CPI developed a method for continuously updating the sample of rented housing units by replacing one-sixth of the rented housing unit sample every year based on the latest available Census Bureau data. In addition to reducing the age of the sample, the proposed process reduces sample attrition, reflects new construction and changes in where people live more accurately, and makes the CPI Housing Survey more efficient.

Since rents are not as volatile as most other consumer prices, it is more efficient for the Housing Survey to have a large sample collected less frequently. To accomplish this, each selected segment in a Pricing Area is assigned to one of six panels, each of which is a representative sub-sample and provides sufficient information for monthly Housing indexes. One panel is priced each month, and each panel is priced twice a year; panel 1 is collected in January and July, panel 2 in February and August, and so on. Every month BLS field staff collects the rent and other information for one-sixth of the sample units. The CPI computes the six-month price ratio (the current rent divided by the rent six months ago) of units in the sample. The measures of price change for the two CPI Housing components are based on averages of these rent ratios.

As samples age and become out of date, they become smaller. This is certainly true of the Housing Survey's sample. Over time, some renter-occupied units become owner-occupied, while others become non-residential. As the burden on respondents accumulates, and for some becomes excessive, there are more refusals. More importantly, because there is no opportunity for newly built rental units to enter the sample, the sample become less representative of the housing universe. Rent change for newer units may be different from that of older ones. By July 2009, when the efforts to improve the sample began, the sample that was introduced in 1999 had

16,000 fewer rental housing units than its target size. See Appendix 1 for details on **Housing Sample Selection**.

Sample replacement and sample augmentation are ways to counteract these problems. Sample replacement drops the old sample items and brings in new ones, selected from a new, more recent sampling frame. Augmentation supplements an existing sample, by adding new sample items drawn from either the original sampling frame or a new frame. An augmentation that uses a new frame requires care to assure consistency of the old and new sample observations, particularly with respect to the sample weights.

The geographic sample of the last CPI Revision, which deployed in January 1998, partitioned the urban United States into 38 Index Areas and selected 87 Pricing Areas from the 1990 Census to represent these areas. These areas were metropolitan areas or urban places and were selected using probability-proportional-to-size (PPS), where the size measure was the 1990 population. For the Housing Survey, the Pricing Areas were further partitioned into small geographic areas called segments, formed from one (in most cases) or more Census block groups and containing at least 50 housing units in large (A-size) Pricing Areas and at least 30 in smaller (B/C-size and D-size) Pricing Areas. A sample of segments was selected in each area using PPS, where the size measure was the sum of renters' actual rents and owners' estimated implicit rents, which the Census provides by block group. About five rental housing units were selected within each segment.

New Housing samples

Most of the rental units in the current (early 2012) Housing Survey entered the sample in the first six months of 1999 when the CPI completely replaced the sample with a new one drawn using the 1990 Decennial Census. Some units built after the 1990 Census were included through a special new-construction survey that covered units built as recently as 2006.

The CPI is undertaking a three-stage effort to improve the Housing Survey within the 87 Pricing Areas selected for the 1998 CPI Revision. The first and second stages used the 2000 Census. The first stage is a two-year sample augmentation. Its goal is to add 16,000 units, mainly in seriously depleted neighborhoods, and bring the size of the sample up to its target. The CPI began using data from this augmentation in the index for July 2010.

The second stage is a four-year sample replacement that will replace the rental units introduced in 1999. The November 2012 CPI will be the first that uses new sample from this stage. The May 2016 CPI will be the first in which the Housing sample will have been drawn entirely from the 2000 Census.

The final stage will be a regular replacement commencing in 2016 and ending in 2022. It will replace the 2000-Census-based sample with one based on either the 2010 Decennial Census or the American Community Survey (ACS). This will continue into the future and - for the first time—the CPI Housing Survey will have a process that keeps its sample continuously updated.

Because of lessons learned in the 1999 sample's selection and initiation, the CPI is using two important new survey methods in the augmentation and replacements. First, commercially available address lists are being used in place of a manual listing process, resulting in significant

cost savings. Second, a pre-screening survey sent out by mail is being employed to identify and screen out owner-occupied housing units and non-residential units in advance of in-person data collection efforts. (Owner-occupied units are out of scope for the Housing sample.)

Previously to draw the Housing Survey sample, BLS statisticians created special maps that indicated the selected segments and blocks of the Pricing Areas. Using these maps CPI field staff travelled to selected segments and “listed” them, meaning that they recorded the addresses of all housing units on their computers. (They use tablet computers for all data collection activities.) Then their computers used a sampling algorithm to select a subset of these addresses. The field staff then visited each of these addresses and found a respondent to interview. They determined whether the address was eligible for the Housing Survey and, if so, initiated those addresses into the Housing Survey. In many segments, a majority of the addresses was screened out because they were non-residential or owner-occupied or failed some other requirement.

Creating maps and listing entire segments are very expensive processes. To reduce these costs, the 2010 sample augmentation used purchased address lists and a mail prescreening survey to locate housing units in the segments. These steps greatly increased the chances that an address contacted by a BLS agent would be a renter-occupied housing unit eligible for the Housing survey. The lists indicate the probability that an address is owner-occupied and the addresses provide a means of determining whether an address is for a commercial establishment. This information was used to determine sampling rates for the prescreening survey. This is a mail survey of the selected addresses to determine if they are commercial or residential and, if they are residential, their tenure (owner or renter occupied). Only those addresses the survey identified as renter-occupied and those with no response are sent out for screening by BLS agents. Preliminary results were successful, so purchased address lists and prescreening will likely be used in all future sample replacements.

BLS field staff must find an eligible respondent for each address and then use tablet computers with Computer-Assisted Data Collection (CADC) software for all data collection. During screening, the software directs them through a structured series of questions. They verify that the unit is renter-occupied. They further determine that the unit is the primary residence of the occupant, that the occupant is not a relative of the landlord, and that the unit is not institutional or public housing, nor an assisted living facility with Activities of Daily Living (ADLs) provided to an occupant. These determine if a selected address is eligible for the Housing sample. Once the field staff determines a unit is eligible, they initiate it into the survey.

The field staff has a multi-month period to screen and initiate the units in a segment. Those units that the field staffs are not able to screen (usually because they fail to contact a respondent) go back out “on panel” for another screening attempt. This process should yield an expected number (usually 5) of in-scope housing units in each segment that will be initiated into the Housing sample.

Initiation and Pricing

Once a selected address has been successfully screened, the CPI field agent immediately proceeds to initiate the housing unit into the Housing sample. Initiation is the initial collection of rent data, which the field agent obtains by asking another structured series of questions. These data include the rent that is paid and specific housing services that are associated with the unit.

These data are the basis for all calculations of rent change that will occur during the life of the housing unit in the Housing sample.

Once a unit is initiated, it will be priced on panel every six months. In addition, any in-scope units that are not successfully initiated go back out on panel for another attempted initiation. A housing unit's initiation generally does not take place in its on-panel month, so the housing unit must be priced on panel for two cycles to provide a six-month interval before price changes can be used in the CPI's estimates of price changes.

Collected data for each unit in the Housing Survey

During initiation and during each pricing, BLS collects:

- Contract rent and rental period (monthly, bi-monthly, weekly or for a specified number of days)
- Utilities, facilities and any other such items included in the rent
- Any subsidies (e.g., Section 8) or reductions in the rent in exchange for services the tenant provided
- Any extra charges included in the contract rent for optional items such as parking
- The number of rooms, type of housing structure, and other physical characteristics
- Equipment used for air conditioning (A/C) and fuels used for heat and hot water.

In addition, the screening questions are asked every two years, or whenever there is a change of occupant, to ensure that the unit remains in scope.

Derived data for each unit in the Housing Survey

After housing data come in from the field, Housing Survey software computes or “derives” additional data elements for each unit. Derived data include review flags and the value of utilities and certain facilities used for quality adjustment. See Appendix 2 for details on **Quality Adjustments**.

An important derived data element is the *normalized rent* of each unit. (Normalized means put on a monthly basis.) Normalized rent is the contract rent plus any subsidies the landlord receives and the value of any work that the tenant provides—all put on a monthly basis, if not so already. Put simply, normalized rent is the price of the “housing service” the unit provided its occupant during the collection month. See Appendix 3 for details on how the normalized rent is derived from the collected data.

However, the service a housing unit provides changes from one collection month to the next one six months later. At a minimum, the housing unit is six months older. Research has shown that as a housing unit gets older the value of housing service that it provides declines. Of course, there may be less subtle changes. For example, the landlord may stop providing heat or electricity, or may have made physical changes to the unit. Because the service a housing unit provides changes between collection periods, the CPI cannot estimate rent change from the change in the normalized rents as calculated in two different collection periods. The key CPI requirement for the prices it uses to calculate price change is that the prices be for comparable items (goods or services) in two periods.

For these reasons the Housing system calculates two prices each month for *Residential rent* index calculation, and, because its underlying concept is different, another two prices for

Owners' equivalent rent index calculation. Economic rent is the CPI price for the *Residential rent* index calculation. The Housing system computes the current-period economic rent and the six-month-previous economic rent for each housing unit each time it is on-panel. Similarly, “pure rent” is the CPI price for the *Owners' equivalent rent* index calculation. The Housing system computes its current-period pure rent and its six-month-previous pure rent for each housing unit each time it is on-panel.

The *Residential rent* index uses a contract rent concept in which items the landlord provides at no extra cost (e.g., utilities) are part of the service the housing unit provides. (The rent index must make a quality adjustment if the extras change, but as long as they do not change they are part of the economic rent.) The concept for *Owners' equivalent rent* is defined as what an owner-occupied housing unit would rent for unfurnished and without utilities. Pure rent, therefore, excludes the estimated value of any landlord-provided utilities. See Appendix 3 for details on how the current economic and pure rents are derived from the current period normalized rent and how the six-month-previous economic and pure rents are derived from the economic and pure rents that were computed six-month-previous.

Housing Index Calculation

Every month the Housing system calculates a one-month relative of price change for each of the 38 CPI Index Areas for the two Housing series (*Residential rent* and *Owners' equivalent rent*) and hands these off to the CPI Index system for calculation of basic and higher-level aggregate indexes. Each one-month relative of price change is the sixth-root of a six-month relative of price change.

A six-month Housing relative is the ratio of the weighted average current-period prices of the rental units in the Index Area's housing sample to the weighted average six-month-previous prices of the same rental units. For the *Residential rent* averages, the current and previous prices of a sampled housing unit are its current and previous economic rents, of course. Its weight is the average number of dollars of rental expenditure the sampled housing unit represents in the renter-occupied housing universe. For the *Owners' equivalent rent* averages, the current and previous prices of a sampled housing unit are its current and previous pure rents and its weight is the number of dollars of implicit rental value a sampled housing unit represents in the owner-occupied housing universe. See Appendix 4 for details on **Housing Index Calculation**.

Appendix 1: Housing Sample Selection

To achieve a geographic distribution of segments within the Pricing Areas, the 2000 Census-based sample first sorted segments by geography (state, county, and Census tract), by their average rent level and then selected a sample of segments by applying systematic PPS sampling, with the cost of housing in a segment as the measure of size. The cost of housing in a segment is the cost of rented housing in the segment plus the cost (implicit rent) of owned housing in the segment. This process guaranteed that the sample for each Pricing Area included segments from different geographic neighborhoods and rent levels.

Segments are one or more block groups (BGs). The 2000 Decennial Census provided the numbers of renters (R_{BG}) and owners (O_{BG}) and the average rent of renter units (RR_{BG}) for each block group. BLS estimated the average implicit rent of the owner units (IR_{BG}) in the block

groups using a regression model for each Index Area based on Consumer Expenditure Survey (CE) data:

$$\begin{aligned}
 IR_{BG} = b_0 &+ b_1 * \text{Average Property Value of Owners}_{BG} \\
 &+ b_2 * (\text{Average Property Value of Owners}_{BG})^2 \\
 &+ b_3 * \text{Median Income of Owners}_{BG's Tract} \\
 &+ b_4 * \text{Average Number of Rooms for Owners}_{BG}
 \end{aligned}$$

The actual regression coefficients (b_0, b_1, b_2, b_3 and b_4) were determined uniquely within each Index Area. The model was chosen for four major reasons:

1. The model is a relatively simple linear regression.
2. The three independent variables (i.e., property value, owner income and number of rooms) are variables expected to affect OER.
3. The model does not depend on any renter information. Therefore, no assumptions about renter-owner interactions have to be theorized.
4. These three CE variables have comparable housing variables in the Census 2000 Summary File 3 (SF3). This comparability allowed the BLS to use the SF3 files as the data source.

From these four pieces of information, the CPI calculated the total cost of housing (TC_{BG}) in the block groups from the renter costs (RC_{BG}) and the owner costs (OC_{BG}) in the block groups.

$$TC_{BG} = RC_{BG} + OC_{BG} = (R_{BG} * RR_{BG}) + (O_{BG} * IR_{BG})$$

Since segments are one or more block groups (BGs):

$$OC_s = \sum_{BG \in s} OC_{BG} \quad RC_s = \sum_{BG \in s} RC_{BG} \quad TC_s = \sum_{BG \in s} TC_{BG}$$

Each segment within each Pricing Area has a probability of selection (P_s) that is the ratio of the cost of housing (TC_s) in the segment multiplied by the number of segments to be selected in the Pricing Area ($n_{PricingArea}$) divided by the total housing cost in the Pricing Area.

$$P_s = \frac{TC_s * n_{PricingArea}}{\sum_{S \in PSU} TC_s}$$

The segment weight is the inverse of the probability of selection,

$$W_s = \frac{\sum_{S \in PricingArea} TC_s}{TC_s * n_{PricingArea}}$$

where 'S' is a segment in a Pricing Area.

Appendix 2: Quality Adjustments

Most of the quality adjustments made to the economic and pure rents are derived automatically. The adjustments mentioned above are described below:

1. **Age bias adjustment** – The quality of all units deteriorates over time. Units get older between observations, so the BLS adjusts the rent for the estimated loss in quality due to aging. A hedonic regression estimates aging effects (among other things). Age bias adjustments are the only adjustments that are applied to the previous rents to reduce the previous pure and economic rents and increase the price change to the current period. The annual impact is an increase of about 0.3 percent in the RENT and OER indexes.
2. **Structural change adjustment** - The hedonic regression mentioned above is also used to derive factors to adjust for any changes in the numbers of bedrooms, bathrooms and other rooms and the conversion to or from Central A/C. If a bedroom, bathroom and/or other room is added to the unit, or if either central A/C is added or non-central A/C is converted to central A/C, the adjustments are subtracted from the current rent to make it consistent with the previous rent. If a bedroom, bathroom and/or other room is removed from the unit, or if either central A/C is removed or the central A/C is converted to non-central A/C, the adjustments are added to the current rent to make it consistent with the previous rent.
3. **Removal of the cost of landlord-provided utilities (Cost of Utilities) for the calculation of Pure Rents**
 - A. **Electricity, Fuel oil and Gas** - Regression coefficients for energy consumption are derived using data from the Department of Energy’s Residential Energy Consumption Survey (RECS). The characteristics of the housing units and the RECS regression coefficients are used to estimate the consumption of electricity, fuel oil and gas for each housing unit. Average prices from the C&S survey are used to derive costs for the three fuels for each housing unit. If the landlord provides the utility, the cost is added to the cost of utilities for the unit.
 - B. **Water and Sewer** - If the landlord provides water and/or sewer services, dollar adjustments are added to the cost of utilities for the unit. Combined water and sewer amounts for each Pricing Area were derived from 2010 ACS data. C&S data were used to divide the combined water and sewer amounts into separate water and sewer amounts at the Pricing Area level. Index data were used to update the water and sewer amounts from 2010 to a current collection period. Water and sewer amounts are also adjusted monthly using index changes. Water and sewer amounts will be updated annually when the ACS data for the following year are available.
4. **Utility Adjustment for Economic Rents** - If the landlord changes the provision of utilities from the previous period to the current period, the economic rent is affected. The individual utility costs derived earlier in the process are used to derive the utility adjustments. If the landlord starts providing electricity, fuel oil, gas, water and/or sewer services, the utility adjustments are subtracted from the current rent to make it consistent with the previous rent. If the landlord stops providing electricity, fuel oil, gas, water and/or sewer services, the utility adjustments are added to the current rent to make it consistent with the previous rent.
5. **Facility adjustments for parking and A/C units**
 - A. **Parking** - If the landlord starts or stops providing free parking, there is a dollar adjustment at the Pricing Area level. If the landlord starts providing free parking, the parking adjustment is subtracted from the current rent to make it consistent with the previous rent. If the landlord stops providing free parking, the parking adjustment is

added to the current rent to make it consistent with the previous rent. Parking amounts are adjusted monthly using index changes.

B. **A/C Units** - If the landlord changes the number of window or through-the-wall A/C units or heat pumps, or starts or stops providing them, there is a cost adjustment. If the landlord starts providing units or increases the number of units, the A/C unit adjustment is subtracted from the current rent to make it consistent with the previous rent. If the landlord stops providing units or decreases the number of units, the A/C unit adjustment is added to the current rent to make it consistent with the previous rent. The A/C cost adjustment is re-evaluated every year. If the A/C equipment changes from window or through the wall to Central A/C or vice versa, the Central A/C structural change adjustment handles the change.

6. **Analyst adjustments** – Rarely one or more of the automated adjustments does not work correctly, in this case the commodity analysts adjust the current and/or previous economic and/or pure rents.

Appendix 3: Calculation of Normalized, Economic and Pure Rents

The normalized rent is the value of what the landlord receives on a monthly basis. The economic rents are the normalized rents adjusted for changes in quality. In most cases, quality adjustments are applied in the current period to make the current and previous period rents consistent. If the landlord provides utilities, the costs of those utilities are embedded in the contract rents, the normalized rents and the economic rents. Economic rents are used in the calculation of the RENT price relatives.

Economic Rents in Summary

Current Economic Rent
= Normalized Rent
+/- Utility Adjustment
+/- Facility Adjustment
+/- Structural Change Adjustment
+/- Analyst Adjustment
T-6 Economic Rent
= Current Economic Rent in T-6
- Age Bias Adjustment
+/- Analyst Adjustment

Pure Rents (PR)

The pure rents also start with the normalized rents and then adjust for changes in the quality of the housing unit, but an additional adjustment is made. Since owners pay their own utilities and the C&S expenditure weight for utilities includes the utility expenses both of owners and of those tenants who pay their own utilities, the value of landlord-provided utilities (Cost of Utilities) must be removed from the normalized rent. Since the Cost of Utilities is subtracted to derive the pure rents, utility adjustments are not applied. Pure rents are used in the calculation of the OER price relatives.

Current Pure Rent

- = Normalized Rent
- Cost of Utilities
- +/- Facility Adjustment
- +/- Structural Change Adjustment
- +/- Analyst Adjustment

T-6 Pure Rent

- = Current Pure Rent in T-6
- Age Bias Adjustment
- +/- Analyst Adjustment

Appendix 4: Housing Index Calculation

Unit Weights

The renter and owner costs in the segments are the basis for the renter and owner weights in the segments used in the Price Relative Calculation.

Renter Weight for the Price Relative Calculation

To derive the renter weight in the segment (RW_s), the segment weight (W_s) must be multiplied by the number of renters in the segment (R_s).

$$RW_s = W_s * R_s$$

The renter weight for a unit in a segment is the renter weight divided by the desired number of renter units in the segment (n_s):

$$RW_{s,i} = W_s * \frac{R_s}{n_s}$$

Owner Weight for the Price Relative Calculation

To derive the owner weight in the segment (OW_s), the segment weight (W_s) must be multiplied by the number of owners in the segment (O_s) and adjusted by the ratio of the estimated average implicit rent of the owner units in the segment (IR_s) to the average rent in the segment (RR_s).

$$OW_s = W_s * O_s * \frac{IR_s}{RR_s}$$

The owner weight for a unit in a segment is the owner weight divided by the desired number of renter units in the segment (n_s):

$$OW_{s,i} = W_s * \frac{O_s}{n_s} * \frac{IR_s}{RR_s}$$

The renter and owner weights are ratios of expenditures, not expenditures themselves, so there is no need to convert them into quantities by dividing them by base rents or base implicit rents.

Price Relative Calculation for RENT and OER

Due to the panel structure of the housing data, 6-month relatives are calculated.

- 6-month relatives for RENT for Index Area (A) are calculated using renter weights ($RW_{S,i}$) and economic rents for the units (i) in the segment in the current month ($ER_{S,i,t}$) and 6 months previous ($ER_{S,i,t-6}$).
- 6-month relatives for OER for Index Area (A) are calculated using owner weights ($OW_{S,i}$) and pure rents for the units (i) in the segment in the current month ($PR_{S,i,t}$) and 6 months previous ($PR_{S,i,t-6}$).

$$\text{Rel}_{A,t-6,t}^{\text{Rent}} = \frac{\sum_{i \in S} RW_{S,i} * ER_{S,i,t}}{\sum_{i \in S} RW_{S,i} * ER_{S,i,t-6}} \quad \text{Rel}_{A,t-6,t}^{\text{OER}} = \frac{\sum_{i \in A} OW_{S,i} * PR_{S,i,t}}{\sum_{i \in A} OW_{S,i} * PR_{S,i,t-6}}$$

Housing takes the sixth root of the 6-month relatives to calculate 1-month relatives and passes them to the Estimation System.

$$\text{Rel}_{A,t-1,t}^{\text{OER}} = \sqrt[6]{\text{Rel}_{A,t-6,t}^{\text{OER}}} \quad \text{Rel}_{A,t-1,t}^{\text{Rent}} = \sqrt[6]{\text{Rel}_{A,t-6,t}^{\text{Rent}}}$$

Calculation of Indexes for RENT and OER

The Estimation System uses the 1-month relatives to move the indexes from last month (t-1) into the current month (t).

$$\text{IX}_{A,t}^{\text{OER}} = \text{IX}_{A,t-1}^{\text{OER}} * \text{Rel}_{A,t-1,t}^{\text{OER}} \quad \text{IX}_{A,t}^{\text{Rent}} = \text{IX}_{A,t-1}^{\text{Rent}} * \text{Rel}_{A,t-1,t}^{\text{Rent}}$$