Weighting for CINCH And Rental Dynamics Analysis: Logic of the Weighting And Final Algorithms

February 2005

Econometrica, Inc.
under contract to:

U.S. Department of Housing
and Urban Development
*Office of Policy Development
and Research*

Principal Authors: Frederick J. Eggers & Fouad Moumen

WEIGHTING FOR CINCH and RENTAL DYNAMICS ANALYSIS: Logic of the Weighting and Final Algorithms

This paper presents a conceptual overview of the weighting problem, suggests a solution, and proposes algorithms for implementing the weighting.

The paper draws upon the appendixes in the most recent CINCH reports,[1] the appendix to the rental dynamics study by Nelson-Vandenbroucke,[2] the weighting appendix to the followup rental dynamics study by Watson-Eggers,[3] copies of some of the correspondence between Gregory Watson of ICF Consulting and Dennis Schwanz of the Census Bureau related to the weighting for the most recent CINCH reports, and conservations with Gregory Watson.

The paper blends ideas from all these sources and attempts to create a conceptually correct and practicable weighting scheme that is roughly equivalent to the scheme used in the most recent CINCH reports. The paper notes some differences between the proposed approach and other approaches. This final version of the paper incorporates comments from David A. Vandenbroucke and Gregory Watson on an earlier draft.

Terminology

Previous CINCH analysis has distinguished between the *status* of a unit with respect to the housing stock—for example, existing as a nonresidential structure—and the *characteristics* of the unit or its occupants—for example, rental versus owner-occupied or the race of the householder. We will use this same distinction. Also adopting previous CINCH terminology, we will refer to 2002 as the *current* year and the previous AHS survey year, either 1994 or 1995, as the *base* year. Finally, observations in the current year and base year with the same control number are *same* if and only if the unit existed in both years and the Census Bureau was able to conduct interviews in both years. If the unit was not newly constructed between the base year and the current year and if observations are not *same*, then the observations are *nonsame*. Observations linked to newly constructed units are neither *same* nor *nonsame*.

Same units provide a physical connection between the base year and the current year. We can look at a unit's status and characteristics in both years and see how they may have changed. Because the final weights assigned to units by the Census Bureau will change

---

[1] *Components of Inventory Change:1993-1995*, a report prepared by ICF Consulting for the U.S. Department of Housing and Urban Development, available on the HUDUSER website, June 2003.
[2] Kathryn P. Nelson and David A. Vandenbroucke, *Affordable Rental Housing: Lost, Stolen, or Strayed?*, a paper presented at the 1996 mid-year meeting of the American Real Estate and Urban Economic Association, May 28, 1996.
[3] *Rental Market Dynamics: Is Affordable Housing For The Poor An Endangered Species?*, a study prepared by ICF Consulting and Econometrica, Inc. for the U.S. Department of Housing and Urban Development, available on the HUDUSER website, December 31, 2003.

between the base year and the current year, *same* units will account for a different number of units in the 2 years. In the tracking process, *same* units will not account for all of the source or destination units. Newly constructed units will add to the current year stock. *Nonsame* units account for the remaining source or destination units. The approach suggested in this paper and the approach used in the previous CINCH reports use only some of the *nonsame*s.

Weighting Logic

It would be possible to list for every unit its status and characteristics in both the base year and the current year. In some cases, there may be no status—for example, not yet constructed in the base year—or no characteristics—for example, no race of householder for vacant units; but, with this understanding, such a listing would still be possible. From the listing, one could construct an exact accounting of the movement of units among the various statuses and characteristics between the base year and the current year.

The exact accounting would apply only to AHS sample observations, roughly a 1-in-2200 picture of the housing stock at the national level. To obtain estimates of the magnitude of actual housing stock changes in status or characteristic, one needs to apply weights to the sampled units. When weights are applied, the accounting will no longer be exact because units have different weights in different years.[4] For example, the exact accounting might show that 400 sample units in a given metropolitan area that were rental in the base year became owner-occupied in the current year. To estimate the number of units in the metropolitan housing stock that were rental in the base year and became owner-occupied in the current year, one would need to apply weights. But using the base year weights will produce a different estimate than using the current year weights. There is no conceptual reason to favor the answer using base year weights over the answer using current year weights. The choice of weights depends upon how the intended analysis will be used. (The impact of weighting is actually much more complicated and intractable.)

For this reason, previous CINCH analysis has distinguished between:

>   (A) *Forward looking analysis*, that is, starting with the base year stock and
>   determining the status and characteristics of *those* units in the current year. The
>   goal here is to explain what happened to the *x* units comprising the housing stock
>   in the base year. Forward looking analysis takes the housing stock as given in the
>   base year and looks at the destination of these units in the current year.

---

[4] The Census Bureau assigns both a pure weight (the inverse of the probability of selection) and a final weight to each AHS observation. The final weights are designed to sum up to independent estimates of the total housing stock. The pure weights will vary over observations within a given AHS because of stratification in drawing the sample. The pure weight of a given observation will vary between surveys if the sample size changes. (Some limited variation in pure weights of given units has been observed over time independent of any change in sample size. These are generally errors.) The final weights will differ over observations within a given AHS because the Census Bureau makes adjustments for various factors affecting the sample. The final weights of a given observation will vary between AHS surveys because of changes in the housing stock.

(B) *Backward looking analysis*, that is, starting from the current year stock and determining the status and characteristics of *those* units in the base year. The goal here is to explain where the *y* units comprising the current year housing stock came from. Backward looking analysis takes the current year housing stock as given and looks at the source of these units, either in the base year or in new construction.

We will follow the same procedure.

In the discussion below, we use examples to explain the CINCH problem and illustrate one conceptually correct solution. We initially abstract from certain complications in order to understand the logic behind weighting. After explaining the logic, we will discuss the complications and how to deal with them conceptually.

Background to the Examples

To understand CINCH weighting, one must distinguish between what happens physically to units and what happens in the interview process. In the examples, we track units both physically and in terms of interviews. The Census Bureau classifies noninterviews into three categories:[5]

- *Type A noninterviews* include cases in which the Census Bureau could not contact the respondent or the respondent refused to be interviewed.

- *Type B noninterviews* consist of units that are temporarily out of the housing stock. Reasons include nonresidential use or being uninhabitable or subject to a condemnation order.

- *Type C noninterviews* indicate units that are permanently out of the housing stock, including demolitions and disaster losses.

Type A noninterviews create difficulties for CINCH weighting.

The examples involve two hypothetical metropolitan areas: one with 200,000 housing units in the base year and one with 200,000 housing units in the current year.[6] We also assume a 1-in-2000 sampling rate. At this rate, we would be tracking 100 sampled units forward from the base year and 100 sampled units backwards from the current year. In the diagrams on the following page:

μ=units	I=interview	A=Type A noninterview
B=Type B loss	C=Type C loss	S=*Same*
NS=*nonsame*	NC=new construction	

---

[5] The Appendix lists the various types of Type A, Type B, and Type C noninterviews.
[6] The examples treat two *different* metropolitan areas. No attempt was made to trace forward and backward between the same metropolitan area.

These diagrams present the two examples. The top diagram illustrates forward looking analysis. The arrows trace forward, at both the physical and interview level, where base units went in the current year. The bottom diagram illustrates backward looking analysis. The arrows trace backwards, at both the physical and interview level, where current year units originated. At the interview level, the diagrams differentiate between *same* and *nonsame* units; the *nonsames* come from two sources: units that were not interviewed for any reason in the non-reference year and Type A noninterviews in the reference year.
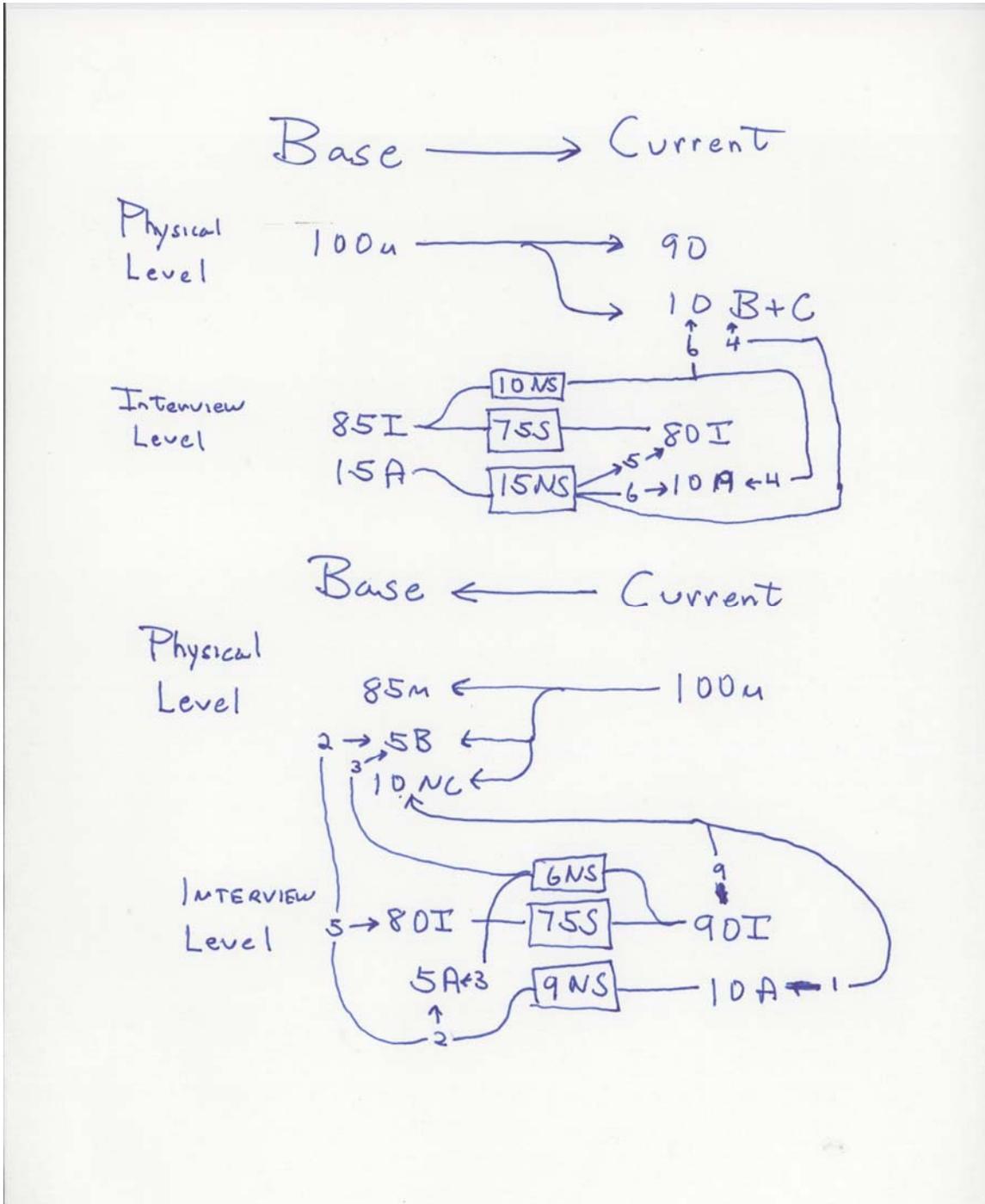
Forward Looking: From Base Year to Current Year

At the physical level, there were 100 sampled units in the base year. In the current year, 90 of the original 100 are still in the housing stock, 10 are either Type B or Type C losses. Note that the actual housing stock in the current year may be more than 90 units because of new construction or the return to the housing stock of units that were Type B losses in the base year. However, we are not interested in these units for the forward looking analysis because they were not part of the base year stock.

At the interview level, the Census Bureau interviewed 85 of the 100 units in the base year; there were 15 type A noninterviews. In the current year, the Census Bureau interviewed 80 of the 90 surviving units; there were 10 type A noninterviews.

Of the 85 units interviewed in the base year and the 80 units interviewed in the current year, 75 were *same* units. Since new construction does not enter into forward looking analysis, the number of *nonsames* equals the total sample minus the number of *sames;* that is, 100-75 = 25. Of the 85 units interviewed in the base year, 10 were not interviewed in the current year. These account for 10 of the *nonsames*; 6 were type B or type C losses and 4 were type A noninterviews in the current year. The remaining 15 *nonsames* originate from the 15 type A noninterviews in the base year. Of these 15, 5 account for the extra interviews in the current year; that is, the 80 current year interviews consist of 75 *sames* and 5 *nonsames.* The 10 remaining base year type A noninterviews were split between type B and C losses (4) and type A noninterviews in the current year (6).

In the base year, the weighting problem can be approached in different ways. We have 85 interviewed units to use to describe 200,000 housing stock units. The final weights assigned to the 85 units by the Census Bureau will sum to 200,000, so we could simply use their final weights. Alternatively, we could break the 85 into component parts, the 75 *same* units and the 10 *nonsame* units and adjust their pure weights (or their final weights) accordingly. We will follow this second approach. In the previous CINCH analysis, the four *nonsame* units that were type A noninterviews in the current year were dropped from the analysis. We will do the same.

The second approach for describing the base year characteristics of the housing stock involves two steps. First, we must adjust the pure weights of the 75 same units.

An appropriate adjustment is to multiply every pure weight by:

$$\frac{\text{Total number of base year units} - \text{total number of type B \& C losses}}{\text{Sum of the base year pure weights for the } \textit{sames}}$$

Total number of base year units = 200,000
Total number of type B&C losses = 10*2000 = 20,000[7]
Sum of base year pure weights for the *sames* = 75*2000 = 150,000

$$\frac{200,000 - 20,000}{150,000} \quad = \quad \frac{180,000}{150,000} \quad = \quad 1.2$$

Next, we have 10 sampled units at the physical level to describe the base year characteristics of the 20,000 current year housing stock units that became type B or C losses. Because they are type B or C losses, these units were not interviewed in the current year and do not have current year characteristics. We are interested only in their base year characteristics. Of the 10 sampled units, only 6 were interviewed in the base year. These 6 units have to describe the base year characteristics of the 20,000 current year losses. Again we must adjust the pure weights of these 6 sampled units.

An appropriate adjustment is:

$$\frac{\text{Sum of the base year pure weights of all type B \& C losses}}{\text{Sum of base year pure weights of interviewed units that became B\&C losses}}$$

Sum of the base year pure weights of all type B&C losses = 10*2000 = 20,000
Sum of base year pure weights of interviewed units that became B&C losses = 6*2000 = 12,000

$$\frac{20,000}{12,000} \quad = \quad 1.67$$

In the current year, we have 90 sampled units that survived from the base year. Using base year pure weights, this represents 180,000 housing stock units. We have 80 sampled unit interviews to describe the 180,000 current year housing stock units that survived from the base year. Seventy-five of the 80 are *sames;* that is, the sample units existed and were interviewed in the base year. We have another 5 units that were interviewed in the current year but were type A noninterviews in the base year.

We could either use or not use the 5 units that were type A noninterviews in the base year. In the previous CINCH analysis, these units were dropped from the analysis. We will do the same. Therefore, we must adjust the pure weights of the 75 same units.

---

[7] This implicitly assumes that the sample was drawn in the base year so that the pure weights sum up to the base year housing stock. This will not be true in the planned analysis. We adjust for this in the proposed algorithm.

An appropriate adjustment is:

$$\frac{\text{Total number of base year units} - \text{total number of type B \& C losses}}{\text{Sum of the base year pure weights for the sames}}$$

Total number of base year units = 200,000
Total number of type B&C losses = 10*2000 = 20,000
Sum of base year pure weights for the *sames* = 75*2000 = 150,000

$$\frac{200,000 - 20,000}{150,000} \quad = \quad \frac{180,000}{150,000} \quad = \quad 1.2$$

This is the same adjustment used for the *sames* regarding base year characteristics, which seems to be a desirable feature.

Backward Looking: How Current Year Came from Base Year

At the physical level, there were 100 sampled units in the current year. Of these, 85 existed in the base year; 5 units were added by the return to the housing stock of type B losses; and 10 units were added by new construction.

At the interview level, 90 of the 100 sampled units were interviewed in the current year; 10 were type A noninterviews. In the base year, 80 of the 85 units that were part of the base year stock were interviewed; the remaining 5 were type A noninterviews.

Of the 90 units interviewed in the current year and the 80 units interviewed in the base year, 75 were same units. There were 100 sample units in the current year. Of these, 10 were newly constructed and 75 were *sames*. The total of *nonsames* was 100-10-75=15. Of the 90 units interviewed in the current year, 15 were not interviewed in the base year. Of these 15, 9 were newly constructed units that had not existed in the base year. The remaining 6 were *nonsames*; 3 were type B losses in the base year and 3 were type A noninterviews in the base year. The remaining 9 *nonsames* were type A noninterviews in the current year. There were a total of 10 current year type A noninterviews, 9 *nonsames,* and 1 that was a new constructed unit.

The current year housing stock was 200,000.[8] To derive housing stock characteristics in the current year, we can focus on the component parts. But first we need to estimate the size of the components.

---

[8] At this point, the examples become more complex because the pure weights cannot be related to the current year stock, that is, the sum of the pure weights are not equal to the current year stock. The Census Bureau could not have drawn a sample in the current year and then gone back in time to sample the same units in the base year. We will deal with this complication in working through the backward looking example. In the examples, we assume the samples were drawn in the base year. In the proposed analysis, the samples were drawn prior to the base year, a further complication that we will deal with later.

In the sample, 10 units come from new construction. We will assume that the Census Bureau drew the 10 units using the same 1-in-2000 sampling rate assumed for the base year sample. However, we will also assume that it is necessary to increase the weight to 2100 in order to represent their share in the current year stock. Therefore, 21,000 units came from new construction (10*2100).

In the sample, 5 units came from the return of units that had been type B losses in the base year. These units have a pure weight of 2,000; we will also assume that it is necessary to increase the weight to 2,200 in order to represent their share in the current year stock. Therefore, 11,000 came from the return to the housing stock of type B losses (5*2200).

By subtraction, the remaining 168,000 units came from units that existed in the base year.

Of the 100 sampled units in the current year, 90 were interviewed. Of these 90, 75 were *sames;* that is, interviewed in both years. As we did in the case of the forward looking analysis, we will not consider units not interviewed in both years. This means that the pure weights of the 75 *sames* have to be adjusted. An appropriate adjustment is:

<u>Total units – (total of Type B returns and total of new construction)</u>
Total of *sames*

Total units = 200,000
Total of Type B returns = 11,000
Total of new construction = 21,000
Total (pure weights) of *sames* = 75*2000 = 150,000

$$\frac{200{,}000 - (11{,}000 + 21{,}000)}{150{,}000} = \frac{168{,}000}{150{,}000} = 1.12$$

Using current year final weights, we calculated that 21,000 came from new construction. Of the 10 sampled newly constructed units, only 9 were interviewed; one was a type A noninterview. The Census Bureau does not compute final weights for newly constructed units separately, thus the sum of the final weights across all newly constructed units will not equal the sum of pure weights of the newly constructed units.[9] The current year pure weights for the newly constructed units will be adjusted by the following ratio:

---

[9] The Census Bureau does adjust the weights for newly constructed units to account for known deficiencies in sampling new construction.

<u>Sum of the current year *adjusted* weights for all newly constructed units</u>
Sum of the pure weights for all units that were newly constructed and interviewed

Sum of the current year adjusted weights of all newly constructed units = 21,000
Sum of pure weights for all units that were newly constructed and interviewed = 18,000

$$\frac{21,000}{18,000} \quad = \quad 1.17$$

Using current year adjusted weights, we calculated that 11,000 came from units that were type B noninterviews in the base year. Of the 5 sampled units that were previously type B losses, only 3 were interviewed. The appropriate adjustment is:

<u>Sum of the *adjusted* weight for all units that were previously Type B</u>
Sum of the pure weights for all units that were previously Type B and interviewed

Sum of the adjusted weights for all units that were previously type B = 11,000
Sum of the pure weights for all units that were previously type B and interviewed = 6,000

$$\frac{11,000}{6,000} \quad = \quad 1.83$$

Having constructed the characteristics of the current year stock from the component parts, we can look at the characteristics of current year units that existed in the base year. Using current year final weights, we calculate that 168,000 came from units that existed in the base year. Of the 85 sampled units that existed in the base year, 80 were interviewed. Of these 80, 75 were *sames*; that is, interviewed in both years. As we did in the case of the forward looking analysis, we will not consider units not interviewed in both years. This means that the pure weights of the 75 *sames* have to be adjusted. An appropriate adjustment is:

<u>Total units – (total of type B recoveries and total of new construction)</u>
Total of *sames*

Total units = 200,000
Total of Type B returns = 11,000
Total of new construction = 21,000
Total (pure weights) of *sames* = 75*2000 = 150,000

$$\frac{200,000 - (11,000 + 21,000)}{150,000} \quad = \quad \frac{168,000}{150,000} \quad = \quad 1.12$$

This is the same ratio used for *sames* in estimating characteristics in the current year, which is a desirable feature.

Complications

Types of interviews: There are three different types of interviews: occupied interviews, usual residence elsewhere (URE) interviews, and vacant interviews. Vacant interviews contain less information than occupied interviews because there are no occupants to interview. URE interviews differ slightly from vacant interviews. The Census Bureau treats URE units as vacant units. The analysis treats units as *same* if there were an interview of any type in both years, including URE interviews.

Elimination of panels to reduce survey costs: The Census Bureau organizes AHS data collection for metropolitan surveys into nine panels of observations at each site. HUD chose to eliminate one panel at each site (three panels in Phoenix) in 2002 to reduce survey costs. The approach used in the most recent CINCH report eliminates observations that appear in only one year (except for new construction). Nelson-Vandenbroucke included these observations and adjusted the pure weights accordingly. Conceptually correct weighting procedures can be constructed using either approach. We will drop the excluded panels from the 1994 (or 1995) surveys to be consistent with the previous CINCH reports, which eliminated the extra units in the national survey resulting from the rural oversample and the metropolitan oversample for the six largest metropolitan areas.

Note that in the forward looking analysis we could have also chosen to use the 5 current year interviews that were type A noninterviews in the base year, and that in the backward looking analysis we could also have chosen to use the 5 base year interviews that were type A noninterviews in the current year. If we had selected this option, we would have had to derive different adjustment factors.

Variation in pure weights: Occasionally the pure weight of an observation will vary from one survey year to another. To guard against this, we will use the maximum of the pure weight in the 2 years. Using the maximum pure weight also helps correct for the effects of dropping a panel or panels.[10]

Census Bureau adjustments contained in the final weights: In simple terms, final weights result from adjustments that the Census Bureau makes to the pure weights so that the interviewed sample sum up when weighted to independent estimates of the housing stock. If this were actually the case, then one could simply ratio up the pure weights to produce the desired counts. In reality, final weights incorporate a more complex set of adjustments that the Census Bureau believes are necessary to present an accurate description of the housing stock. Using only the pure weights loses the fine tuning that the Census Bureau applies in producing final weights. For this reason, we tried to construct the algorithms around the final weights, but were unable to arrive at a workable approach because some of the units that we want to follow from one period to another have zero final weights. For example, in the forward looking

---

[10] If units were in the sample at the time of both the base year and the current year surveys and if the units were only interviewed in one of these years, we drop the units from the analysis. For this reason, we need to use the larger of the two pure weights.

analysis, we want to follow all units that existed in the base year but were type B or type C losses in the current year. Some of these units were type A noninterviews in the base year, and thus have zero final weights in that year. We need weights for these units in order to compute the total number of type B and type C losses. Therefore, we were forced to rely on ratio adjustments to the pure weights.

Algorithms

We will discuss algorithms for both the forward looking and the backward looking analyses. The goal of our approach is to produce forward looking estimates that sum to the AHS totals, by MSA, of the base year (either 1994 or 1995) housing stock, and backward looking estimates that sum to the AHS totals, by MSA, of the current year (2002) housing stock.

To achieve this goal, we follow a three-step procedure:

- Calculate an appropriate CINCH weight for each observation.
- Determine which observations to eliminate.
- Determine which observations need to have the CINCH weights adjusted and calculate those adjustments.

We apply the algorithms to each MSA separately.

Forward Looking: From Base Year to Current Year

Here are the steps necessary to prepare the data to analyze what happened between the base year and the current year to units that existed in the base year. AHS variables are given their codebook names and presented in capital letters. We refer to base year variables by the suffix IN1994_. For the five sites surveyed in 1995, this should be IN1995_. Current year variables are labeled IN2002_.

1. Merge each 2002 metropolitan file with the corresponding 1994 or 1995 file. Use the 2002 flat file. Keep non-matches—this is important because of the dropped panels in the 2002 surveys. HUD and the Census Bureau redesigned the AHS questionnaire in 1997, and thus there are differences between the pre-1997 surveys and surveys conducted in 1997 and later. For this reason, adjustments will have to be made for differences in variables, variable names, and variable coding between surveys. The following steps are performed separately for each metropolitan area.

2. For all units let MXPWT = max (IN2002_PWT, IN1994_PWT). In almost all cases this will be the IN2002_PWT because of the dropped panels in the 2002 survey year. If there were an unusually low IN2002_PWT for one or more observations, this step would adjust for that low value. (PWT is the pure weight.)

3. Compute an estimate of the base year stock (BASECOUNT) by summing IN1994_WEIGHT across all observations. (WEIGHT is the final weight, not the pure weight.)

4. Eliminate from subsequent analysis:

   a. All observations that have no matches—this will eliminate the units that were in the panels dropped from the 2002 surveys. (It will also eliminate new construction but this is of no consequence for forward looking analysis since newly constructed units were not part of the base year housing stock.)

   b. All observations that were base year type B or type C losses (10 GE IN1994_NOINT LE 38).[11] These units were not part of the base year stock and therefore are not tracked in the forward looking analysis.

   c. All observations that were deleted in the base year prelist subsampling (IN1994_NOINT=39).

5. Compute SMXPWT = sum of MXPWT after step 4; this sum is a first estimate of the size of the housing stock based on the units retained for analysis.

6. Compute a FLCINCHWT = MXPWT*(BASECOUNT/SMXPWT). This computation ratios the weights up so that they sum to the base year stock.

7. Identify *sames*, *losses*, and *interview losses*:[12]

   a. SAME=1 if IN1994_ISTATUS = 1,2, OR 3 AND IN2002_ISTATUS = 1,2,OR3.

   b. LOSS = 1 if IN1994_ISTATUS=1,2,3,OR4 AND 10 GE IN2002_NOINT LE 38.

   c. INTLOSS = 1 if IN1994_ISTATUS=1,2,OR3 AND 10 GE IN2002_NOINT LE 38.

8. Calculate:

   a. SSAMES = sum of FLCINCHWT for all SAME = 1.
   b. SLOSSES = sum of FLCINCHWT for all LOSS = 1.
   c. SINTLOSSES = sum of FLCINCHWT for INTLOSS = 1.

---

[11] See the appendix for an explanation of these codes.
[12] ISTATUS = 1 (regular interview) ISTATUS = 2 (Usual Residence Elsewhere interview) ISTATUS = 3 (vacant interview) ISTATUS = 4 (noninterview).

9. Eliminate from subsequent analysis all observations that were base year or current year type A noninterviews. We cannot use the noninterviews because there is no information on the characteristics of these units. However, we retain them until this point so that we can get good estimates of the number of losses (SLOSSES).

10. Calculate:

   a. Ratio1 = (BASECOUNT – SLOSSES)/SSAMES.
   b. Ratio2 = SLOSSES/SINTLOSSES.

11. Recalculate FLCINCHWT as follows:

   a. For SAMES =1, FLCINCHWT = Ratio1*FLCINCHWT.[13]
   b. For INTLOSS =1, FLCINCHWT = Ratio2*FLCINCHWT.

12. From published reports obtain estimated base year counts for all occupied units, all vacant units, and all seasonal units. Calculate:

   a. Sum of FLCINCHWT for all SAME = 1 or INTLOSS = 1 in which ISTATUS = 1 (occupied units). Adjust the FLCINCHWT for these observations so that they sum to the published total for occupied units.

   b. Sum of FLCINCHWT for all SAME = 1 or INTLOSS = 1 in which 2 LE ISTATUS LE 3 AND NOT(8 LE VACANCY LE 10) (URE and vacant units).[14] Adjust the FLCINCHWT for these observations so that they sum to the published total for vacant units.

   c. Sum of FLCINCHWT for all SAME = 1 or INTLOSS = 1 in which (IN02_ISTATUS='2' OR IN02_ISTATUS='3') AND (8 LE IN02_VACANCY LE 10) (Seasonal units). Adjust the FLCINCHWT for these observations so that they sum to the published total for seasonal units.

---

[13] This is programming code, not an algebraic equation. The code replaces the old value of FLCINCHWT that appears on the right-hand side of = with the value equal to Ratio1 times the old value.

[14] VACANCY is the AHS variable for type of vacancy. Codes 8, 9, and 10 identify seasonal-summer only, seasonal-winter only, and other seasonal units respectively.

Backward Looking: From Current Year to Base Year

Here are the steps necessary to prepare the data to analyze where current year units came from. AHS variables are given their codebook names and presented in capital letters. Current year variables are labeled IN2002_. We refer to base year variables by the suffix IN1994_. For the five sites surveyed in 1995, this should be IN1995_.

1. Merge each 2002 metropolitan file with the corresponding 1994 or 1995 file. Use the 2002 flat file. Keep non-matches—this is important because of the dropped panels in the 2002 surveys and units constructed between the base year and the current year. HUD and the Census Bureau redesigned the AHS questionnaire in 1997, and thus there are differences between the pre-1997 surveys and surveys conducted in 1997 and later. For this reason, adjustments will have to be made for differences in variables, variable names, and variable coding between surveys. The following steps are performed separately for each metropolitan area.

2. For all units let MXPWT = max (IN2002_PWT, IN1994_PWT). In almost all cases this will be the IN2002_PWT because of the dropped panels in the survey year. If there were an unusually low IN2002_PWT for one or more observations, this step would adjust for that low value. (PWT is the pure weight.)

3. Compute an estimate of the current year stock (CURRENTCOUNT) by summing IN2002_WEIGHT across all observations. (WEIGHT is the final weight, not the pure weight.)

4. Eliminate from the sample:

    a. All units found in the base year survey but not in the current year survey. This will eliminate the panels dropped from the 2002 survey. (Note that we are keeping new construction; that is, units found in the current year survey but not found in the base year.)

    b. All type B or type C losses in the current year (10 GE IN2002_NOINT LE 38). These units are not part of the current year stock and therefore we do not track them backwards.

    c. All observations that were deleted in the current year prelist subsampling (IN2002_NOINT=39).

5. Compute SMXPWT = sum of MXPWT after step 4; this sum is a first estimate of the size of the housing stock based on units retained for analysis.

6. Compute a BLCINCHWT = MXPWT*(CURRENTCOUNT/SMXPWT). This computation ratios the weights up so that they sum to the current year stock.

7. Identify *sames*, *recoveries*, *interviewed recoveries, new construction,* and *interviewed new construction*:

    a. SAME=1 if IN1994_ISTATUS = 1,2, OR 3 AND IN2002_ISTATUS = 1,2,OR3.

    b. REC = 1 if IN2002_ISTATUS=1,2,3,OR4 AND 10 GE IN1994_NOINT LE 17.

    c. INTREC = 1 if IN2002_ISTATUS=1,2,OR3 AND 10 GE IN1994_NOINT LE 17.

    d. NC = 1 if IN2002_BUILT GE 1994 (or 1995 as appropriate) AND IN1994_ISTATUS = 99 OR . OR B.

    e. INTNC = 1 IF NC=1 AND IN2002_ISTATUS=1, 2, OR 3.

8. Calculate:

    a. SSAME = sum of BLCINCHWT for all SAME = 1.
    b. SREC = sum of BLCINCHWT for REC =1.
    c. SINTREC= sum of BLCINCHWT for INTREC = 1.
    d. SNC = sum of BLCINCHWT for NC =1.
    e. SINTNC = sum of BLCINCHWT for INTNC=1.

9. Eliminate from subsequent analysis all observations that were base year or current year type A noninterviews. We cannot use the noninterviews because there is no information on the characteristics of these units. However, we retain them until this point so that we can get good estimates of the number of recoveries (SREC) and the number of newly constructed units (SNC).

10. Calculate:

    a. Ratio1 = (CURRENTCOUNT – (SREC + SNC))/SSAME.
    b. Ratio2 = SNC/SINTNC.
    c. Ratio3 = SREC/SINTREC.

11. Recalculate BLCINCHWT as follows:

    a. For SAME = 1, BLCINCHWT = Ratio1*BLCINCHWT.
    b. For INTNC= 1, BLCINCHWT = Ratio2*BLCINCHWT.
    c. For INTREC = 1, BLCINCHWT = Ratio3*BLCINCHWT.

12. From published reports obtain estimated current year counts for all occupied units, all vacant units, and all seasonal units. Calculate:

a. Sum of BLCINCHWT for all SAME = 1 or INTLOSS = 1 in which
    ISTATUS = 1 (occupied units). Adjust the BLCINCHWT for these
    observations so that they sum to the published total for occupied units.

b. Sum of BLCINCHWT for all SAME = 1 or INTLOSS = 1 in which 2 LE
    ISTATUS LE 3 AND NOT(8 LE VACANCY LE 10) (URE and vacant
    units). Adjust the BLCINCHWT for these observations so that they sum to
    the published total for vacant units.

c. Sum of BLCINCHWT for all SAME = 1 or INTLOSS = 1 in which
    (IN02_ISTATUS='2' OR IN02_ISTATUS='3') AND (8 LE
    IN02_VACANCY LE 10) (Seasonal units). Adjust the BLCINCHWT for
    these observations so that they sum to the published total for seasonal
    units.

Appendix – Types of Non-Interviews

The AHS variable that indicates the reason it was not possible to conduct an interview is NOINT. The following is the coding for NOINT. The codes distinguish between Type A, Type B, and Type C noninterviews.

NOINT = Noninterview reason
1	Type A - No one home
2	Type A - Temporarily absent
3	Type A - Refused
4	Type A - Unable to locate
5	Type A - Language problem
6	Type A - Other, occupied
10	Type B - Permit granted, construction not started
11	Type B - Under construction, not ready
12	Type B - Permanent or temporary business or commercial storage
13	Type B - Unoccupied site for mobile home or tent
14	Type B - OTHER unit or converted to institutional unit
15	Type B - Occupancy prohibited
16	Type B - Interior exposed to the elements
17	Type B - not classified above
30	Type C - Demolished or disaster loss
31	Type C - House or mobile home moved
32	Type C - Unit eliminated in structural conversion
33	Type C - Merged not in current sample
36	Type C - Permit abandoned
37	Type C - not classified above
38	Type C - Unit eliminated in subsampling
39	Unit deleted in prelisting subsampling
B	Not applicable