

American Housing Survey

**WEIGHTING STRATEGY FOR
2004 METROPOLITAN CINCH
AND RENTAL DYNAMICS
ANALYSIS**

July 2006

Econometrica, Inc.,
and ICF Consulting
under contract to:
U.S. Department of Housing
and Urban Development
*Office of Policy Development
and Research*

Principal Author: Frederick J. Eggers

WEIGHTING STRATEGY FOR 2004 METROPOLITAN CINCH AND RENTAL DYNAMICS ANALYSIS

HUD has requested Econometrica Inc. and its partner, ICF Consulting, to conduct a components of inventory change (CINCH) analysis and a rental dynamics analysis for the 13 metropolitan areas surveyed in 2004 by the American Housing Survey (AHS). The AHS last surveyed these areas in 1995 or 1996. Table 1 lists the areas, their survey dates, and initial estimates of the sample sizes in both years. In general, the previous surveys have significantly smaller sample sizes because HUD found it necessary to drop survey panels to same money in 1995 and 1996.

Table 1: 2004 Metropolitan Areas with Information on Samples

SMSA Number	Description	2004 Sample Size	Previous Survey	Previous Sample Size	Notes on Previous Sample
520	Atlanta, GA	5,132	1996	4,640	Panels 5, 7, 9, & 11 dropped
1680	Cleveland, OH	4,722	1996	4,602	
2080	Denver, CO	4,834	1995	4,229	Panels 5, 7, 9, & 11 dropped
3280	Hartford, CT	4,728	1996	4,531	
3480	Indianapolis, IN	4,814	1996	4,642	
4920	Memphis, TN-AR-MS	4,644	1996	4,534	
5560	New Orleans, LA	4,516	1995	3,914	Panel 11 dropped
5880	Oklahoma City, OK	4,829	1996	4,471	
6280	Pittsburgh, PA	4,723	1995	4,320	1/2 Panel 11 dropped
6920	Sacramento, CA	4,728	1996	3,983	Panels 5, 7, 9, & 11 dropped
7040	Saint Louis, MO-IL	4,741	1996	4,553	
7240	San Antonio, TX	4,863	1995	4,198	Panel 11 dropped
7600	Seattle, WA	4,731	1996	4,682	Panels 5 & 9 dropped and 1/2 of 11

This paper adapts the weighting strategy used in previous studies for a CINCH analysis of changes in the national housing stock between 1995 or 1996 and 2004.¹ For those new to CINCH analysis, Appendix A contains an overview of the CINCH weighting problem.

One interesting feature of the planned analysis is that it will use as benchmarks for the beginning and ending years published Census Bureau counts based on two different censuses. The 1995 and 1996 weights will be calibrated to counts based on information from the 1990 census while the 2004 weights will be calibrated to counts based on information from the 2000 census. Some of the observed changes will result from differences in the underlying bases.

¹ See in particular: *Weighting Strategy for 2001-2003 CINCH Analysis*, report to HUD by ICF Consulting and Econometrica, Inc., August 2005

Algorithms for Computing Weights

Forward Looking: From 1995 or 1996 to 2004

Here are the steps necessary to prepare the data to analyze what happened between 1995 or 1996 and 2004 to units that existed in 1995 or 1996. AHS variables are given their codebook names and presented in capital letters. We refer to 1995 or 1996 variables by the suffix IN95_ or IN96_; 2004 variables are labeled IN04_. In describing each step, the paper uses IN95_ when IN96_ would be appropriate for the 1996 sites.

1. Merge the 1995 or 1996 and 2004 files, using the flat files. Eliminate cases that appear in the 2004 files but not the 1995 or 1996 files -- this is important because some panels were dropped from the 1995 and 1996 surveys. It will also eliminate new construction and other additions to the stock but this is of no consequence for forward looking analysis since newly constructed units and other additions were not part of the 1995 or 1996 housing stock.
2. For all units let $MXPWT = \max (IN04_PWT, IN95_PWT)$. (PWT is the pure weight). In general, the pure weights should not differ across surveys; however, the PWT in 1995 and 1996 should be higher than the PWT in 2004 because of the dropped panels in those years. If there were an unusually low IN95_PWT for one or more observations, this step would adjust for that low value.
3. Compute an estimate of the 1995 or 1996 stock (BASECOUNT) by summing IN95_WEIGHT across all observations. (WEIGHT is the final weight based on 1980 census geography.)
4. Eliminate from subsequent analysis:
 - a. All observations that were 1995 or 1996 type B or type C losses (10 LE IN95_NOINT LE 38). These units were not part of the 1995 or 1996 stock and therefore are not tracked in the forward looking analysis.
 - b. All observations that were deleted in the 1995 or 1996 prelist subsampling (IN95_NOINT=39).
5. Compute $SMXPWT = \text{sum of } MXPWT \text{ after step 4}$; this sum is a first estimate of the size of the housing stock based on the units retained for analysis.
6. Compute a $FLCINCHWT = MXPWT * (\text{BASECOUNT} / SMXPWT)$. This computation ratios the weights up so that they sum to the 1995 or 1996 stock.
7. Identify *sames*, *losses*, and *interviewed losses*:
 - a. $SAME=1$ if $IN95_ISTATUS = 1, 2, \text{ or } 3$ AND $IN04_ISTATUS = 1, 2, \text{ OR } 3$

- b. $LOSS = 1$ if $IN95_ISTATUS=1, 2, 3, \text{ or } 4$ AND $10 \leq IN04_NOINT \leq 38$
 - c. $INTLOSS = 1$ if $IN95_ISTATUS=1, 2, \text{ or } 3$ AND $10 \leq IN04_NOINT \leq 38$
8. Calculate:
- a. $SSAME = \text{sum of FLCINCHWT for all SAME} = 1$
 - b. $SLOSS = \text{sum of FLCINCHWT for all LOSS} = 1$
 - c. $SINTLOSS = \text{sum of FLCINCHWT for INTLOSS} = 1$
9. Eliminate from subsequent analysis all observations that were 1995 or 1996 or 2004 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 (SLOSS).
10. Calculate:
- a. $Ratio1 = (BASECOUNT - SLOSS)/SSAME$
 - b. $Ratio2 = SLOSS/SINTLOSS$
11. Recalculate FLCINCHWT as follows:
- a. For $SAME = 1$, $FLCINCHWT = Ratio1 * FLCINCHWT$
 - b. For $INTLOSS = 1$, $FLCINCHWT = Ratio2 * FLCINCHWT$
12. From published reports obtain estimated 1995 or 1996 counts for all owner-occupied units, all renter-occupied units, all vacant units, and all seasonal units.²
- Calculate:
- a. Sum FLCINCHWT for all $SAME = 1$ or $INTLOSS = 1$ in which $IN95_ISTATUS = '1'$ (occupied units) AND $IN95_TENURE = 1$ (owner-

² At this point, the algorithm differs from that used in the previous metropolitan CINCH studies (1994 or 1995 with 2002). In the earlier metropolitan studies, we controlled to occupied, vacant, and seasonal units but not for owner-occupied or renter-occupied units. We expanded the controls in the CINCH study using the 2001 and 2003 national surveys because of the importance of tenure in the tables on income and housing costs.

occupied units). Ratio adjust the FLCINCHWT for these observations so that they sum to the published total for owner-occupied units.

- b. Sum FLCINCHWT for all SAME = 1 or INTLOSS = 1 in which IN95_ISTATUS = '1' (occupied units) AND (2 LE IN95_TENURE LE 3) (renter-occupied units). Ratio adjust the FLCINCHWT for these observations so that they sum to the published total for renter-occupied units.
- c. Sum FLCINCHWT for all SAME = 1 or INTLOSS = 1 in which (IN95_ISTATUS='2' OR IN95_ISTATUS='3') AND NOT (8 LE IN95_VACANCY LE 10) (URE and vacant units). Ratio adjust the FLCINCHWT for these observations so that they sum to the published total for vacant units.
- d. Sum FLCINCHWT for all SAME = 1 or INTLOSS = 1 in which (IN95_ISTATUS='2' OR IN95_ISTATUS='3') AND (8 LE IN95_VACANCY LE 10) (Seasonal units). Ratio adjust the FLCINCHWT for these observations so that they sum to the published total for seasonal units.

Backward Looking: From 2004 to 1995 or 1996

Here are the steps necessary to prepare the data to analyze where 2004 units came from. AHS variables are given their codebook names and presented in capital letters. 2004 variables are labeled IN04_; we refer to 1995 or 1996 variables by the suffix IN95_ or IN96_. In describing each step, the paper uses IN95_ when IN96_ would be appropriate for the 1996 sites.

1. Merge the 1995 or 1996 and 2004 files, using the flat files. Keep non-matches—this is important because units are added to the AHS sample between surveys.
2. For all units let $MPWT = \max(\text{IN04_PWT}, \text{IN95_PWT})$. (PWT is the pure weight). In general, the pure weights should not differ across surveys but the pure weights should be higher for surveyed cases in years with dropped panels. If there were also an unusually low IN04_PWT for one or more observations for other reasons, this step would also adjust for that low value.
3. Compute an estimate of the 2004 stock (CURRENTCOUNT) by summing IN04_WEIGHT across all observations. (WEIGHT is the final weight, not the pure weight.)
4. Eliminate from the sample:

- a. All units found in the 2004 survey but not in the 1995 or 1996 survey except units added to the stock since 1995 or 1996. To do this we will drop all units in the 2004 survey that were not in the 1995 or 1996 surveys except for units with IN04_REUAD values. This will eliminate the units that were contained in the dropped panels.
 - b. All type B or type C losses in the 2004 (10 LE IN04_NOINT LE 38). These units are not part of the 2004 stock and therefore we do not track them backwards.
 - c. All observations that were deleted in the 2004 prelist subsampling (IN04_NOINT=39).
5. Compute $SMXPWT = \text{sum of } MXPWT$ after step 4; this sum is a first estimate of the size of the 2004 housing stock based on units retained for analysis.
6. Compute a $BLCINCHWT = MXPWT * (CURRENTCOUNT / SMXPWT)$. This computation ratios the weights up or down so that they sum to the 2004 stock.
7. Identify *sames*, *new construction*, *interviewed new construction*, *other adds*, and *interviewed other adds*:³
- a. $SAME = 1$ if IN95_ISTATUS = 1, 2, or 3 AND IN04_ISTATUS = 1, 2, OR 3
 - b. $NC = 1$ if IN04_ISTATUS=1, 2, 3, or 4 AND ((IN04_REUAD = 3) OR (10 LE IN95_NOINT LE 11))
 - c. $INTNC = 1$ IF $NC=1$ AND IN04_ISTATUS=1, 2, or 3
 - d. $ADD = 1$ if IN04_ISTATUS=1, 2, 3, or 4 AND ((4 LE IN04_REUAD LE 11) OR (12 LE IN95_NOINT LE 17))
 - e. $INTADD = 1$ if $ADD = 1$ AND IN04_ISTATUS=1, 2, OR 3
8. Calculate:
- a. $SSAME = \text{sum of } BLCINCHWT$ for all $SAME = 1$
 - b. $SNC = \text{sum of } BLCINCHWT$ for $NC = 1$
 - c. $SINTNC = \text{sum of } BLCINCHWT$ for $INTNC=1$

³ Other adds are units that were type B losses in 1995 or 1996 but are in the 2004 housing stock plus new housing units that are not new construction, such as the conversion to residential use of a warehouse or mobile home move-ins.

- d. $SADD = \text{sum of BLCINCHWT for ADD} = 1$
 - e. $SINTADD = \text{sum of BLCINCHWT for INTADD} = 1$
9. Eliminate from subsequent analysis all observations that were 1995 or 1996 or 2004 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 (SADD) and the number of newly constructed units (SNC).

10. Calculate:

- a. $\text{Ratio1} = (\text{CURRENTCOUNT} - (\text{SADD} + \text{SNC})) / \text{SSAME}$
- b. $\text{Ratio2} = \text{SNC} / \text{SINTNC}$
- c. $\text{Ratio3} = \text{SADD} / \text{SINTADD}$

11. Recalculate BLCINCHWT as follows:

- a. For $\text{SAME} = 1$, $\text{BLCINCHWT} = \text{Ratio1} * \text{BLCINCHWT}$
- b. For $\text{INTNC} = 1$, $\text{BLCINCHWT} = \text{Ratio2} * \text{BLCINCHWT}$
- c. For $\text{INTADD} = 1$, $\text{BLCINCHWT} = \text{Ratio3} * \text{BLCINCHWT}$

12. From published reports obtain estimated 2004 counts for all owner-occupied units, all renter-occupied units, all vacant units, and all seasonal units.⁴

Calculate:

- a. Sum BLCINCHWT for all $\text{SAME} = 1$ OR $\text{INTNC} = 1$ OR $\text{INTADD} = 1$ in which $\text{INO3_ISTATUS} = '1'$ (occupied units) AND $\text{IN04_TENURE} = 1$ (owner-occupied units). Ratio adjust the BLCINCHWT for these observations so that they sum to the published total for owner-occupied units.
- b. Sum BLCINCHWT for all $\text{SAME} = 1$ OR $\text{INTNC} = 1$ OR $\text{INTADD} = 1$ in which $\text{IN04_ISTATUS} = '1'$ (occupied units) AND (2 LE IN04_TENURE LE 3) (renter-occupied units). Ratio adjust the

⁴At this point, the algorithm differs from that used in the previous metropolitan CINCH studies (1994 or 1995 with 2002). In the earlier metropolitan studies, we controlled to occupied, vacant, and seasonal units but not for owner-occupied or renter-occupied units. We expanded the controls in the CINCH study using the 2001 and 2003 national surveys because of the importance of tenure in the tables on income and housing costs.

BLCINCHWT for these observations so that they sum to the published total for renter-occupied units.

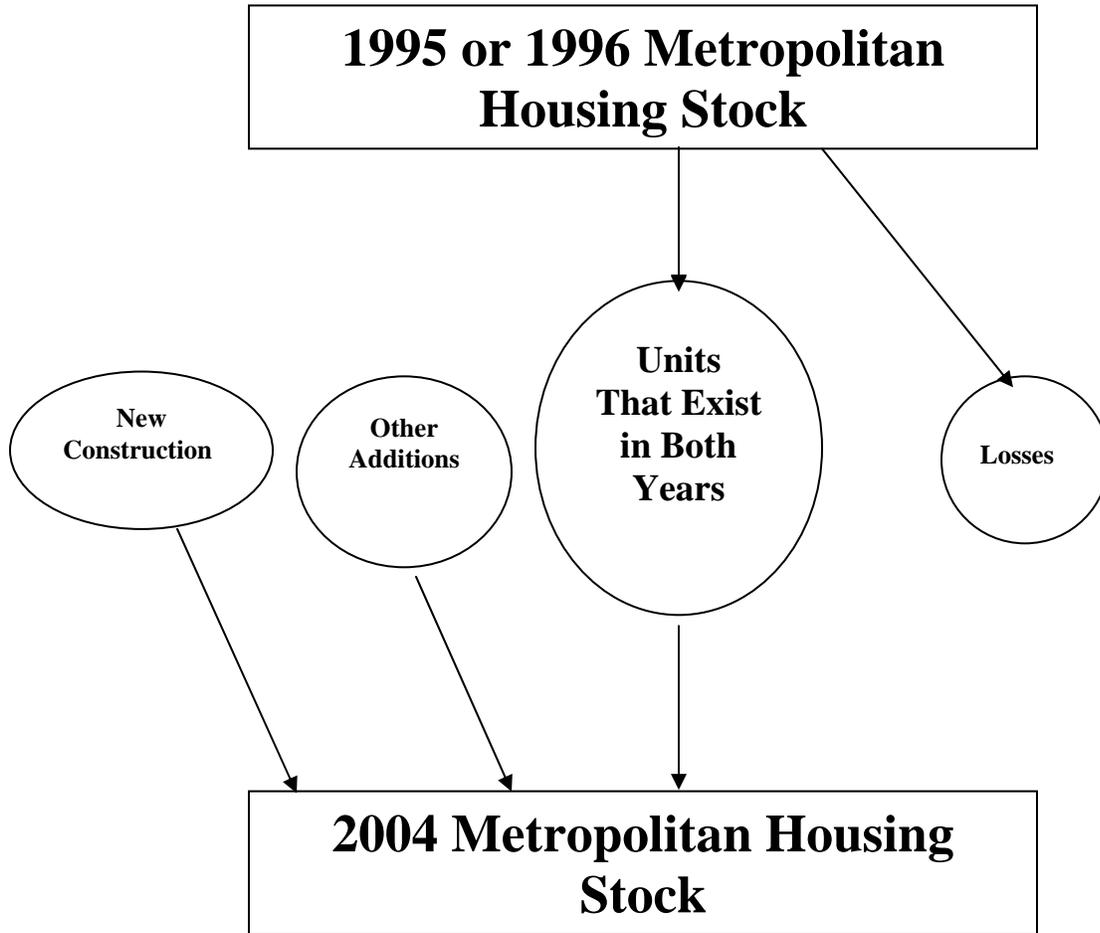
- c. Sum BLCINCHWT for all SAME = 1 OR INTNC = 1 OR INTADD = 1 in which (IN04_ISTATUS='2' OR IN04_ISTATUS='3') AND NOT(8 LE IN04_VACANCY LE 10) (URE and vacant units). Ratio adjust the BLCINCHWT for these observations so that they sum to the published total for vacant units.
- d. Sum BLCINCHWT for all SAME = 1 OR INTNC = 1 OR INTADD = 1 in which (IN04_ISTATUS='2' OR IN04_ISTATUS='3') AND (8 LE IN04_VACANCY LE 10) (Seasonal units). Ratio adjust the BLCINCHWT for these observations so that they sum to the published total for seasonal units.

Appendix A: Overview of CINCH Weighting

The CINCH Objective

Figure 1 illustrates the question that CINCH analysis seeks to answer.

Figure 1: The CINCH Objective



CINCH tries to explain how the housing stock evolves from one period to the next. Figure 1 contains 4 ovals and two rectangles. The Census Bureau provides estimates for both rectangles and one oval (units added through new construction between 1995 or 1996 and 2004). No one estimates the other three ovals: the number of units that belong to both the 1995 or 1996 and 2004 housing stocks, units loss to the housing stock between 1995 or 1996 and 2004, and other additions to the housing stock between 1995 or 1996 and 2004.

Losses can be either permanent or temporary. Units destroyed by natural disasters or intentionally demolished are permanent losses. Temporary losses include units that are

merged into other units or units that are used for non-residential purposes.⁵ Besides new construction, additions can include units resulting from splitting up larger units, mobile home move-ins, and units that had been used formerly for non-residential purposes.

In addition to determining the size of each oval, housing analysts find information about the characteristics of the units in the different ovals useful. Interesting characteristics include: structure type, age of the unit, size of the unit, tenure, household size and composition, resident income, and resident race and ethnicity.⁶

CINCH analysis has three goals:

- To provide estimate for all six components of Figure 1
- To disaggregate losses and other additions into relevant component parts
- To characterize the units that survive from one period to the next and the units that are added or lost between periods.

The AHS has four features that make CINCH analysis possible:

- Each unit has weights that can be used to estimate its share of the overall stock.
- The AHS tracks new construction and the various types of losses and other additions.
- The AHS has detailed information about the characteristics of each unit and its occupants.
- The AHS tracks the same unit from one period to the next so that changes in status and characteristics can be observed directly.

Weighting

Ideally analysts would like to solve two simultaneous equations using CINCH analysis.⁷

- (1) 1995 or 1996 housing stock = units that exist in both years + losses
- (2) new construction + other additions + units that exist in both years = 2004 housing stock

Unfortunately previous experience with CINCH analysis has shown it is difficult to find satisfactory simultaneous solutions to the equations. For this reason, previous research by Econometrica and ICF solved the two equations separately.

In previous studies, solving equation (1) was termed forward-looking analysis because it tracks what happens to the units in the 1995 or 1996 housing stock. In terms of Figure 1, forward-looking analysis deals with the top rectangle and the two ovals on the right.

⁵ “Potentially reversible” might be a better term than “temporary” for these types of losses.

⁶ In CINCH analysis involving the national housing stock, interesting characteristics also include location by region and location by metropolitan status.

⁷ The equations are “simultaneous” because the term “units that exist in both years” appears in each equation.

In previous studies, solving equation (2) was termed backward-looking analysis because it tracks where units in the 2004 housing stock came from. In terms of Figure 1, backward-looking analysis deals with the bottom rectangle and the three ovals on the left. In analytical terms, backward-looking analysis reverses the arrows at the bottom of Figure 1 by taking the 2004 housing stock as its starting point.

Separating the analysis into forward-looking and backward-looking components results in each observation having two weights, a weight for the forward-looking analysis (FLCINCHWT) and a weight for the backward-looking analysis (BLCINCHWT).

Solving the equations separately also results in two independent estimates of “units that exist in both years,” one based on each set of weights. This paper develops algorithms to carry out the forward-looking and backward-looking analyses.