

Effect of Housing Unit Controls on Survey Estimates
Keith Albright
U.S. Census Bureau
Keith Albright, U.S. Census Bureau, Demographic Statistical Methods Division,
Washington, D.C. 20233

Keywords: American Community Survey, Census 2000 Supplementary Survey, Housing Unit Controls, Weighting, Housing Post-Stratification Factor, Principal Person Factor

1. Introduction

Control totals are often used in calculating weights for survey respondents. The control totals are used to calculate a factor which is used to adjust the weights. This has the effect of making the weighted total equal to the control total. The Census Bureau's American Community Survey (ACS) and Census 2000 Supplementary Survey (C2SS) use housing unit (HU) control totals to make the weighted total number of HUs equal to the number of HUs counted in the 2000 Census.

The ACS is a monthly survey that collects demographic and socioeconomic data about households and persons and is intended to replace the decennial census long form. Testing of the ACS began in 1996, and is now in 36 counties grouped into 31 sites. Full implementation of the ACS is scheduled to begin in every county in the United States in 2003. The C2SS used the same methods and instrument as the ACS, surveying approximately 700,000 HUs in 1,203 counties. It was conducted to test the feasibility of conducting a large scale national survey concurrently with the decennial census. The C2SS was designed to be used in conjunction with the ACS to produce national and state level estimates.

The 3,142 counties in the United States were grouped into primary sampling units (PSUs). The PSUs are grouped into strata. A strata may consist of only one PSU, and counties with more than 250,000 people are in a PSU and strata by themselves. Each ACS site is a strata by itself. Strata are classified as self representing (SR) and non-self representing (NSR). All PSUs in SR strata are selected to be in sample. In NSR strata, two PSUs are selected to be in sample. ACS sites are considered SR strata. In weighting, the strata are used to form cells for which adjustment factors are computed.

In 2000, the ACS/C2SS used the count of HUs in the

2000 decennial census for control totals. But in future years, we will not have the advantage of a census conducted in the same year to provide control totals. This paper examines the effect that HU controls have on estimates in the ACS/C2SS. We attempt to determine if HU controls are necessary for producing accurate and reliable estimates. We also examine alternative HU estimates to determine their effectiveness as control totals.

We also examine the use of the principal person factor (PPF) in the weighting of housing units to determine if it is necessary to continue using it.

Housing controls and the PPF are used to compute adjustment factors for the HU weights after the application of nonresponse adjustment factors. The housing controls are used to compute a housing post stratification factor (HPF1) to make the weighted number of HUs equal to the control total. The HU weights are then assigned to each person in the HU. Population controls, from the Census Bureau's population estimates program, are used to compute post stratification factors for each person on the basis of race, age, sex, and Hispanic origin. The post stratification factor of the principal person (the PPF) is then applied to the HU weight. The housing controls are then used to compute a second housing post stratification factor (HPF2) to make the weighted number of HUs again equal to the control total. The final weight is the product of the base weight (determined by the probability of selection) and all of the adjustment factors.

2. Use of Housing Unit Controls and the PPF

2.1 Use of Control Totals

In the 2000 ACS and Supplementary Survey, we used the results of the decennial census to provide HU controls for weighting HUs. This will not be the case in future years. The HU controls are used to calculate a

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau Publications. This report is released to inform interested parties of research and to encourage discussion.

housing post stratification factor (HPF) to adjust the weights. We will examine two options:

- Not using HU controls at all.
- Using the yearly HU estimates calculated by the Census Bureau’s Population Division as control totals (we will refer to these as ‘HU estimates’).

2.2 Use of the Principal Person Factor

The principal person of a household is either the female spouse of the reference person, or the reference person if there is no female spouse. The PPF has the effect of setting the housing unit weight equal to the weight of the principal person. It attempts to correct household weights for bias due to race and Hispanic origin. We will determine if we should keep using this factor in the future.

3. **Weighting Methods**

3.1 Reweightings

We will use seven different reweightings of the 2000 data, along with the original weights, to evaluate the issues in section 2. These reweightings are:

1. Use HU estimates with PPF
2. Exclude HPF, include PPF
3. HPF2 Alternative 1 (described below)
4. HPF2 Alternative 2 (described below)
5. Include HPF, exclude PPF
6. Use HU estimates without PPF
7. No HPF or PPF

Methods 3, 4, and 5 use the census as controls, while 1 and 6 use the annual HU estimates as controls. Methods 2 and 7 use no HU controls.

Throughout this paper, we will refer to these different weighting methods by their number in the above list. The original weighting will be referred to as C2SS or C2SS weighting.

3.2 Alternatives for the Housing Post Stratification Factor (HPF)

If we decide to use HU controls, we may change the way in which we apply them. The HPF is calculated for each stratum as follows:

$$HPF = \frac{\text{Control Count}}{\sum WGTS}$$

The numerator is the number of HUs to which our estimate of total HUs is being controlled (for the 2000 ACS/C2SS, this is the census count). The denominator is the weighted number of HUs at the previous weighting step. So all HUs in a stratum get the same HPF and there is no distinction between occupied and vacant HUs.

There are currently two housing post stratification adjustment factors: HPF1 and HPF2. HPF1 is calculated after the nonresponse adjustments, with a single value being calculated for all HUs in a stratum. Then the PPF is applied, which adjusts the weights of occupied HUs. HPF2 is calculated after the PPF so that the total HU estimate will again equal the control. Currently, a single value of HPF2 is calculated and applied to all HUs in a stratum, even though the weight of vacant HUs did not change after the PPF was applied (by definition, vacant HUs do not have a principal person so their PPF is equal to one). So the weights of vacant HUs are adjusted again due to an adjustment that was only made to occupied HUs.

We are considering two alternative methods for calculation of HPF2

- Calculate HPF2 only for occupied HUs while still controlling to the total number of HUs.
- Calculate HPF2 separately for occupied and vacant HUs.

These alternative methods are only being considered for HPF2. HPF1 will continue to be computed as before.

4. **Analysis**

4.1 Comparison of Estimates

Estimates of the number of housing units will be calculated under each weighting method. We will look at the following estimates:

- Total HUs
- Total HUs by building type
- Total occupied HUs
- Occupied HUs by tenure
- Total vacant HUs
- Vacant HUs by vacancy reason
- Occupied HUs by tenure
- Occupied HUs by race and Hispanic origin of householder

Various comparisons of estimates from these reweightings will be used to evaluate the effects of using HU controls and the PPF. These comparisons are

given below. There are additional comparisons that can be used, but we will limit the number to keep the analysis manageable. We are only evaluating the effects of these methods on weights, not whether or not one method is better than any other. Other analyses will be used to determine which methods are best.

Two sets of comparisons will be used to examine the effect of the census HU controls on estimates. One comparison uses the PPF and the other does not.

- Method 2 and C2SS
- Methods 7 and 5

Two sets of comparisons will be used to examine the effect of the PPF on estimates. One comparison uses HPF and the other does not.

- Method 5 and C2SS
- Methods 7 and 2

Comparing method 1 to the C2SS will allow us to see the joint effect of not using HU controls or a PPF.

Three sets of comparisons will be used to examine the effect of using the HU estimates as control totals.

- Method 1 and C2SS
- Methods 1 and 2
- Methods 6 and 7

The second two comparisons examine the effect of HU estimates compared to estimates obtained when no HU controls are used at all.

Methods 3 and 5 will be compared to C2SS to determine the effect of alternative methods of calculating HPF2.

4.2 Differences in Individual Weights

We will also evaluate the effect of the different weighting methods on the weights of individual HUs. This will be done by computing the difference between the C2SS weight and the new weight, resulting in a set of seven differences for each HU. We will examine descriptive statistics of the differences and their absolute values.

4.3 Standard Errors and Root Mean Square Error

Standard errors (SE) and root mean square errors (RMSE) of estimates from the reweightings and original C2SS weighting will be compared to help determine which method is best. The SEs are calculated by

replication methods (successive differences). The RMSE of an estimate is given by $\sqrt{SE^2 + BIAS^2}$, where BIAS is the bias of an estimate compared to the 2000 Census.

We used a nonparametric method [Miller (1981)] related to the Friedman test that simultaneously tests each pair of weighting methods. The null hypothesis is that there are no differences between treatments (reweightings). As in the Friedman test, the measures of each treatment are ranked from smallest to largest within each block (states). The null hypothesis is not rejected if, for each pair of treatments, the inequality

$$|\bar{R}_i - \bar{R}_{i'}| \leq q_{k,\alpha}^{.05} \left[\frac{k(k+1)}{12n} \right]^{\frac{1}{2}}$$

$$i, i' = 1, \dots, k$$

is satisfied. The constant $q_{k,\alpha}^{.05}$ is the .05 percentile point of the studentized range, \bar{R}_i is the average rank across blocks for the i^{th} treatment, k is the number of treatments, and n is the number of blocks. A difference between treatments i and i' is indicated when the inequality is not satisfied.

States that have no sample cases with the characteristic being tested will be excluded because there is no estimate of standard error. For example, a number of states did not have any sample cases where the householder was Native Hawaiian/Other Pacific Islander.

4.4 Loss Functions

Loss functions will be computed to determine which weighting method is best. The loss function is given by:

$$L = \frac{\sum_i (X_i - C_i)^2}{C_i}$$

where X_i is the estimate for the i^{th} state and C_i is the census count for the i^{th} state. The items that the loss functions will be produced for are the items listed in section 4.1, except for building type which is not available for the census (not a short form item).

5. Results

5.1 Comparison of Estimates

The estimates that are compared are national level estimates. We made pairwise comparisons of the confidence intervals to determine if there were

significant differences. Table 4 shows the estimates for total, occupied, and vacant HUs.

The use of census HU counts as controls has a significant effect on the estimates of total HUs and occupied HUs, compared to what they would be if controls were not used. The direction of the difference depends on whether or not the PPF is used. There is no significant effect on the estimates of vacant HUs. Comparison of method 2 and C2SS (which use the PPF) shows that the estimates of total and occupied HUs are greater when the census HU controls are not used. Comparison of methods 7 and 5 (which do not use the PPF) shows that the estimates of total and occupied HUs are lower when the census HU controls are not used.

The effect of the PPF on HU estimates depends on whether or not HU controls are used. There is no significant difference between estimates when comparing C2SS and method 5 (which use the HPF). However, we will show later that the PPF can have a large effect on individual weights. These differences, both positive and negative, cancel out when looking at totals. There may be large differences between estimates when looking at subsets of housing units. Comparing methods 7 and 2 (which do not use the HPF) shows that using the PPF results in higher estimates for the total and occupied number of HUs (there is no comparison for vacants because the weight for a vacant HU is the same under these two methods).

Using both HU controls and the PPF results in higher estimates of total HUs and occupied HUs than if they were both not used. But these differences are not as large as the ones were described in the previous two sections

The comparisons of method 1 vs. method 2 and method 6 vs. method 7 are analogous to the comparisons that evaluate the effect of census HU controls described in section 5.1 (recall that these methods use the annual HU estimates as controls). The use of these estimates as controls, using the C2SS weighting methods, results in estimates that are not significantly different than estimates from the C2SS.

The two alternative methods of calculating HPF2 result in estimates for total and occupied HUs that are not significantly different from the C2SS estimates. Method 4 results in vacant HU estimates that are significantly lower than C2SS and method 3.

5.2 Differences in Weights

The different weighting methods could produce individual HU weights much different from those in the C2SS, even though estimates of totals were often similar to the C2SS.

Weighting methods that used the PPF generally had absolute differences that were much less variable than the methods that did not use the PPF. When the PPF was used, the regular differences were almost all positive or all negative. This was not the case with methods that did not use the PPF. This suggests that the use of the PPF adds stability to the individual HU weights. Table 1 shows descriptive statistics of the absolute differences.

Based on these findings, we will continue to use the PPF in housing unit weighting. Therefore, our analyses in the sections which follow only focus on the weighting methods that do not use the PPF. These methods are C2SS weighting, method 1, method 2, method 3, and method 4.

5.3 Standard Errors and Root Mean Square Error

As stated in the previous section, five weighting methods will be compared. This results in 10 treatment pairs for the nonparametric analysis described in the analysis section. Only one characteristic (vacant, for migrant workers) out of the 20 that we examined showed no significant differences among the weighting methods.

In general, there were more significant differences between methods for SEs than for RMSE. The results of the tests indicate that method 4 produces the best estimates in terms of SE and RMSE since it consistently did better than the other methods in its comparisons. However, this weighting method is not a viable method at this time as we will explain later. So we concluded that we should continue to use the C2SS weighting methodology and use the annual HU estimates as control totals. The C2SS weighting and method 1 consistently did better than the other methods they were compared to (except as noted above). And when these two methods were compared, there were almost no significant differences. Table 3 shows the results of all the comparisons.

5.4 Loss Functions

The loss functions do not show one weighting method being uniformly better than the others. For many items, there is not a lot of difference between the loss functions for each method. Method 4 often had the lowest value for the loss function, especially in the

categories of vacant units. This was not surprising since this method controlled both occupied and vacant totals. Loss function values are shown in Table 2.

6. Conclusions

Based on our analysis we made the following decisions:

- We will continue to use the principal person factor. This factor adds stability to individual HU weights.
- We will use the HU estimates provided by the Census Bureau's Population Division as HU control totals. The resulting weights produce estimates that are not significantly different than the ones produced by the C2SS.

- We will not change the way that HPF2 is calculated. The current method produces better estimates, as measured by standard error, root mean square error, and loss functions.

- The second HPF2 alternative (controlling both the occupied and vacant unit totals) would probably be a better method to use. But at this time we can not get separate estimates of occupied and vacant HUs to use as control totals. We will reconsider this method if these estimates become available in the future.

References

Miller, Rupert G. *Simultaneous Statistical Inference, Second Edition*. Springer-Verlag.

Table 1. Statistics of Absolute Differences

Methods With PPF	Mean	Std Dev	Range	IQR
Method 1	4.42	8.14	311	4
Method 2	6.85	11.43	451	7
Method 3	1.52	4.82	220	1
Method 4	7.19	28.60	2993	4

Methods Without PPF	Mean	Std Dev	Range	IQR
Method 5	17.90	30.44	2243	19
Method 6	18.64	31.17	2216	20
Method 7	19.22	32.05	2405	21

Table 2. Values of Loss Functions

	C2SS	Method 1	Method 2	Method 3	Method 4
By Vacancy Reason					
Total Vacant Units	182073	187229	259278	258628	na
For Rent	61958	67290	64637	57856	107801
For Sale Only	32939	33919	38887	35802	39779
Rented/Sold, Unoccupied	89022	87688	93680	98547	64284
Seasonal/Recreational/Occasional	258640	266572	339073	316349	89120
For Migrant Workers	14442	14402	14541	14943	19273
Other Vacant	329379	307580	340858	375688	195243
By Tenure					
Total Occupied Units	17176	37304	15139	24912	na
Owned With Mortgage	302502	306952	200756	316852	250696
Owned Free and Clear	471687	452588	580441	459270	522279
Rented for Cash	24027	39304	34861	25532	19583
No Cash Rent	48803	50233	46398	49140	48650
By Race/Hispanic Origin of Householder					
White	47686	76452	92290	49556	49030
Black	26170	27429	18914	28118	19568
American Indian/AK Native	48123	46101	49759	48097	49021
Asian	19368	21713	24830	19451	20716
Native Hawaiian/ Other Pacific Island	34115	34549	34729	34244	34711
Some Other Race	390677	385421	353974	393669	376595
Non-Hispanic	21096	40370	13413	28700	2354
Hispanic	31661	30941	38124	31947	32789

Table 3. Significance Tests for Standard Errors and Root Mean Square Errors

Comparison* Category	C vs 1		C vs 2		C vs 3		C vs 4		1 vs 2		1 vs 3		1 vs 4		2 vs 3		2 vs 4		3 vs 4	
	By Vacancy Reason																			
Total Vacant Units	-	-	C	C	C	C	#	#	1	1	1	1	#	#	-	-	#	#	#	#
For Rent	-	-	-	-	C	-	-	C	1	-	1	-	-	-	-	-	4	-	4	4
For Sale Only	-	-	-	-	C	-	4	-	1	-	1	-	-	-	-	-	4	-	4	-
Rented/Sold, Unoccupied	-	-	-	-	C	-	-	-	-	-	1	-	-	-	2	-	-	-	4	-
Seasonal/Recreational/Occasional	-	-	C	C	C	C	4	4	1	1	1	-	4	4	-	-	4	4	4	4
For Migrant Workers	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Other Vacant	-	-	-	-	C	C	4	-	-	-	1	1	-	-	-	-	4	4	4	4
By Tenure																				
Total Occupied Units	-	-	-	-	C	C	#	#	-	-	1	-	#	#	-	2	#	#	#	#
Owned With Mortgage	-	-	2	2	C	-	4	-	-	2	1	-	4	4	2	2	-	-	4	4
Owned Free and Clear	-	1	C	C	-	3	-	C	1	1	1	1	1	1	-	2	4	4	-	3
Rented for Cash	-	-	C	-	-	-	4	-	1	-	1	-	4	-	3	-	4	4	4	4
No Cash Rent	-	-	C	-	-	-	-	-	1	2	-	-	-	-	3	-	-	-	-	-
By Race/Hispanic Origin of Householder																				
White	-	-	-	-	C	-	4	-	1	-	1	-	4	4	-	-	4	4	4	4
Black	-	-	C	-	-	-	-	-	-	-	-	-	4	4	3	-	4	-	-	4
American Indian/AK Native	-	-	-	-	-	-	-	-	1	-	-	-	-	-	3	-	-	-	-	-
Asian	-	-	C	-	-	-	-	-	1	-	-	-	-	-	3	-	4	-	-	-
Native Hawaiian/Other Pac Island	-	-	C	-	-	-	-	-	1	-	-	-	-	-	3	-	-	-	-	-
Some Other Race	-	-	C	2	-	-	-	-	1	2	-	-	-	-	3	2	-	-	3	-
Non-Hispanic	-	-	-	-	C	C	4	4	1	2	1	-	4	4	-	2	4	4	4	4
Hispanic	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	4	-	-	-

*Under each comparison, the 1st column shows the result of the standard error comparison and the 2nd column shows the result of the RMSE comparison. The entry in the column indicates which method the test favored.

'C' denotes C2SS weighting. Numbers denote the methods defined in section 3.1

'-' indicates no significant difference in the comparison.

'#' indicates that a test is not appropriate since method 4 controlled the total vacant and occupied units to the census.

Table 4. National Estimates of HUs Under Each Weighting Method

	C2SS	Method 7	Method 2	Method 5
Total	115904650*	114743153	117657335	115904644*
Occupied	104733569	103526368	106440580	104532886
Vacant	11171081	11216785	11216755	11371758
	Method 1	Method 6	Method 3	Method 4
Total	115552193*	115547755*	115904645*	115904661*
Occupied	104524314	104260074	104532950	105480157*
Vacant	11027879	11287681	11371695	10424504*

*Indicates that the estimate is controlled.