Weighting and Estimation Research Methodology and Results From the American Community Survey Family Equalization Project

Mark E. Asiala

US Census Bureau
DSSD, CENHQ RM 4H057, 4600 Silver Hill Rd, Washington DC 20233
mark.e.asiala@census.gov

Abstract

Historically the American Community Survey (ACS) has produced inconsistent estimates of households and householders and inconsistent estimates of husbands and wives in married couple households even though logically these estimates should be equal. In the 2005 ACS, the size of these inconsistencies at the national level was approximately 3.7 million more householders than households and approximately 1.8 million more spouses than married-couple households. Likewise, for unmarried-partner households there are approximately 176,000 more unmarried-partners than unmarried-partner households. The cause of these data inconsistencies was rooted in the person weighting methodology used prior to 2006 which was independent of the housing unit weighting and did not consider relationship to the householder. This paper describes the current weighting methodology and changes introduced to reduce these data inconsistencies while having a minimal impact on other estimates and on the variances of the estimates. A three-dimensional raking methodology is used where the marginal control totals are derived from the survey itself rather than an independent source for the first two dimensions that are related to equalizing spouses and householders. Changes in the estimation of housing unit characteristics are also discussed. Empirical results from the implementation of this new methodology are presented based on the 2004 and 2005 ACS data.

Background

The American Community Survey (ACS) is part of the Census Bureau's plans for a re-engineered 2010 Census. The ACS will collect long-form (sample) data on an annual basis in order to produce single and multi-year estimates. Specifically, the 5-year estimates are comparable to the long-form estimates traditionally produced after each decennial census.

The ACS collects information on a wide variety of topics including housing, household, family, and person characteristics. The existing weighting methodology produces two weights to tally characteristics for all domains: a housing unit weight which was used for housing, household, and family characteristics and a person weight which was used for person characteristics including householder characteristics. Prior to the 2006 ACS, the weighting methodology produced these weights semi-independently with no method to try to ensure consistency between the weights. Because of this, some significant data inconsistencies were present in the ACS data prior to 2006. The largest was the inconsistency between the estimate of householders and the estimate of households. Logically, this should be a one-to-one relationship. However, in 2005 the ACS estimate for householders was 114.8 million but the estimate of households was only 111 million. Similarly, there should be a one-to-one relationship between the number of spouses in households and the number of married-couple households. Again in the 2005 ACS, there were 57.1 million

spouses but only 55.2 million married-couple households. These inconsistencies did not exist in the unweighted data so clearly the weighting methodology was responsible for the observed differences in the estimates. In each of the examples given, the first estimate (householders, spouses) was tabulated from the person weights and the second estimate (households) was tabulated from the housing unit weights. These differences were of great concern to data users both external and internal to the Census Bureau.

Previous Attempts to Address the Problem

The first attempt to address the difference between the estimate of householders and households was in 2002. The solution was to use the person weight of the householder to tabulate estimates for households rather than the housing unit weight. This led to consistency between the estimates of householders and households but introduced a new inconsistency between the estimate of households and occupied housing units which should also have a one-to-one relationship. There were also other undesirable impacts on other estimates which led to abandoning this method and returning to using the housing unit weight to tabulate households.

Our second attempt to address these data inconsistencies borrowed from the weighting methodology of the New York City Housing Vacancy Survey (NYCHVS) conducted by the Census Bureau (Navarro et al., 2004). The NYCHVS methodology defines the final person weight of the householder and of the spouse to be equal to the housing unit weight. The person post-stratification adjustment to the independent population controls is amended so that the householder and spouse by default always get an adjustment equal to 1 and the balance of the sample persons must account for the entire difference between the pre-controlled estimates and the population controls. While this achieves the data consistencies desired at the record level, it does so at high cost. By forcing the balance of the sample persons to account for all of the relative coverage difference, the comparatively larger adjustment factors leads to increased variances in the estimates. The method can also lead to negative weights which must be handled through an increased amount of collapsing of the post-stratification cells. This in turn leads to a greater bias in the estimates as the controls are applied to broader cells. We determined that while this method works for the NYCHVS which is concerned primarily with housing units, it was not appropriate for the ACS with its wide range of estimates for housing units and persons and so it, too, was abandoned.

Research Goals

Based on our experience from the early attempts, the primary goal for the family equalization project was more modest than the record-level consistency achieved using the NYCHVS method. We defined our objective to simply reduce but not necessarily eliminate the inconsistencies at the weighting area level which for the ACS is either a county or a collection of counties. By weakening the objective to the weighting area, we could still obtain our overall goal for most of the estimates but the expectation was that we could do so without the large impact on variances we observed in the NYCHVS method.

It was also recognized that if we achieve success in closing the 3.7 million gap between the estimate of householders and households, that additional person weight would be shifted to other members of the household in an uncontrolled fashion. In order to avoid this potential problem we expanded our goals for the project to the following:

1. To reduce the inconsistency between the estimates of householders and households.
2. To reduce the inconsistency between the estimates of spouses and married-couple households.
3. To reduce the inconsistency between the estimates of unmarried partners and unmarried-partner households.
4. To reduce the inconsistency between the estimates of spouses in subfamilies and married-couple subfamilies.
5. To improve the estimate of vacant housing units.
6. To accomplish the above while not causing unintended impacts on other estimates or variances.

The first four goals directly address the data inconsistencies described previously. Goal #5 was added to the project as the result of another concern about the vacancy rate of the ACS. The source of this concern is described in more

detail in the next section. Lastly, the sixth goal was there to explicitly evaluate the impact of the weighting changes on other estimates and to recognize that achieving these data consistencies at the expense of substantial increase in variances or changing other estimates in a manner that we could not support would not be acceptable.

Before describing the methodology developed to accomplish these goals, we give a quick overview of the ACS weighting methodology as it applies to the project.

Methodology Prior to 2006

The weighting methodology for the ACS is a ratio-estimation method involving several steps. We present here a high-level overview with some additional detail given concerning the person weighting and the final steps of the housing unit weighting. For a more detailed description of the existing weighting methodology prior to 2006, see US Census Bureau (2006).

The weighting methodology produces two sets of weights: a housing unit weight and a person weight. The basic steps of the housing unit weighting are as follows:

1. Calculate the base weights defined as the inverse of the probability of selection.
2. Apply the Variation in Monthly Sample adjustment to smooth month to month response variation.
3. Apply a set of non-interview adjustments.
4. Apply a housing-unit post-stratification adjustment to control the estimate of total housing units.
5. Perform the person weighting [see below].
6. Apply the Principal Person adjustment.
7. Reapply the housing-unit controls via a second housing-unit post-stratification adjustment.

The person weighting was a one-step process. The initial person weight was equal to the housing unit weight after step (4) above and then a person post-stratification adjustment was performed after classifying persons into cells based on race, ethnicity, age, and sex. While the intermediate housing unit weight from step (4) is used as an initial person weight, the methodology prior to 2006 had no mechanism in place to ensure the type of consistency between the final housing unit weight and the final person weight to avoid the data inconsistencies described in the Background section. It is this fact that is the principal cause of the data inconsistencies.

One housing unit weight step of note is step (6) where the person weighting informs the housing unit weighting through the principal person adjustment. The principal person of the household is defined to be the female spouse of the householder, if she exists. Otherwise, the principal person is simply defined to be the householder. The principal person adjustment in the housing unit weighting defines the housing unit weight for all occupied housing units to be equal to the weight of the principal person, effectively incorporating the person post-stratification adjustment of the principal person into the housing unit weight. The purpose of this adjustment is to correct for differential household coverage and net household coverage under the assumption that the only manner in which the principal person can be missed is if the entire household is not included in or missed by the survey.

The impact of the principal person adjustment, however, is a net increase in occupied housing units. In addition, the total number of housing units increased and so step (7) is necessary in order to make the estimate of total housing units to once again agree with the independent controls. The impact of the principal person adjustment on the final housing unit weights still increases the estimate of occupied units but it also decreases the vacancy rate. Analysts in our Housing and Household Economics Statistics division were concerned that the result of this weighting step artificially depressed the ACS estimate of the vacancy rate further away from the Census Bureau's official vacancy rate estimate obtained through the Housing Vacancy Survey (a supplement to the Current Population Survey). While there are structural differences between the two surveys, the ACS estimate of 10.4% ($\pm$0.2%) was significantly lower than the official estimate of 12.8% ($\pm$0.1%). Because of this, we added Goal #5 above to see if we could address this issue at the same time we were addressing Goals #1–#4.

Proposed Methodology
The initial proposed methodology involved changes to the person weighting, the housing unit weighting, and also the estimation. All three parts had to work together in order to address all the goals listed including Goal #6 which was to make these changes without having unintended impacts elsewhere on other estimates or variances.

Weighting steps (1)–(4), listed earlier, were not changed in the proposed methodology. All changes to the weighting affected only steps (5)–(7).

Person Weighting Changes. The change to the person weighting is to introduce a raking-ratio estimation procedure by which we can reduce the data inconsistencies described previously. The Census 2000 sample weighting methodology used a similar approach but it also had the benefit of 100% Census counts for detailed demographics by householder/non-householder, spouses, and the like. While the ACS population controls do not include additional detail, we were able to design a more detailed raking matrix that used ACS estimates in the marginals in order to carry out the raking methodology.

The raking matrix is defined with three dimensions. The first dimension carries out the spousal/unmarried-partner equalization. The second dimension carries out the householder/household equalization. The third, and final, dimension applies the same demographic controls in the same fashion as the person post-stratification adjustment did in the original methodology. The geographic scope of the matrix is our weighting area which is defined to be a county or collection of counties. The order of the dimensions reflects the relative importance of each goal. Consistency with the controls has the highest priority and thus is the final step in the raking to ensure exact agreement with the controls as we have had in the past. The second to last step is the householder/household equalization step which was identified as a higher priority over the spousal equalization.

The spouse equalization dimension is defined as having 3 cells:

- Householder with spouse/unmarried-partner present
- Spouse/unmarried-partner of the householder
- Balance of sample persons

The first two cells are constrained to be equal to the estimate of married-couple or unmarried-partner households which is tabulated using the housing unit weight from step (4). The third cell is then constrained to be equal to the total population from the independent control minus the sum of the control totals for the first two cells. In this manner, total population is always controlled to the independent population control.

A second variation of this raking dimension included the married-couple subfamilies. To accomplish this, the subfamily head was added to the first cell and the subfamily spouse was added to the second cell. The constraint for the first and second cells for this variation are the estimate of married-couple or unmarried partner households plus the total number of married-couple subfamilies using the same housing unit weight as above. The third cell is then constrained to the residual of the total population as was calculated earlier.

The householder equalization dimension is defined as having 2 cells:

- Householders
- Non-householders (balance of sample persons)

The first cell is constrained to be equal to the estimate of occupied housing units and hence the estimate of households. The second cell is then constrained to be equal to the total population control minus the marginal control for the first cell.

The demographic dimension is defined as having up to 156 cells per weighting area obtained from multiplying the

number cells for race/ethnicity (6) by age (13) by sex (2). In practice, collapsing of small cells typically makes the actual number of cells fewer than 156.

Each step, or dimension, in the raking is applied in series to make up one iteration, always ending with the demographic dimension. For the research, the raking was repeated for 10 iterations for every weighting area. No additional convergence criterion was used.

Housing Unit Weighting Changes. The first proposed change to the housing unit weighting methodology was to remove steps (6) and (7) from the list given in the earlier section. This was to accomplish two things. The first is that it removed the source of the depressed vacancy rate. The second is that this allows the person weighting to use the final housing unit weight as the initial person weight. This avoids the possibility that additional steps in the housing unit weighting may adversely affect the consistency gained through the raking done in the person weighting.

Estimation Changes. The final change planned was in the estimation of household characteristics. By removing the principal person adjustment from the housing unit weighting methodology, the housing unit weight will no longer contain this adjustment for differential coverage of households which could negatively impact the estimates for characteristics of households particularly when broken out by the race of the householder.

To account for the differential coverage, we planned to use the weight of the householder for tabulating estimates for households rather than using the weight of the housing unit. This would have two impacts. The first is that the tabulated estimates of householders and households would be exactly equal and the tabulated estimates of householders and occupied housing units would be equal to the extent that the raking is successful in the second dimension. The second impact is that using the householder weight would capture the differential household coverage while essentially assuming that the net household undercoverage is zero (because we are trying to equalize it to our initial estimate of occupied housing units). Our expectation was that the difference between occupied housing units and households should be small for tabulated areas which are at the county level or aggregations of counties.

Final Methodology

The final methodology was changed slightly from the proposed methodology outline above because of the discovery of some unintended consequences in our review of the data. The problems occurred in a set of housing unit tables that were crossed by person or household characteristics. This is discussed in more detail in the Results section. The methodology was changed from above to include one additional step in the housing unit weighting. This step makes the weight of an occupied housing unit to be equal to the weight of its householder. The weight of vacant housing units remain unchanged by this step.

This guarantees agreement between estimates of occupied housing units, households, and householders. It also makes the change in the weight used to tabulate estimates of household characteristics unnecessary since both the householder weight and the housing unit weight are equal.

Results

The methodology described above was used to produce two alternative sets of weights using the 2004 ACS data. The first set included the spouses in subfamily equalization and the second set did not. A total of 313 tables were then created using these two sets of weights and we compared the estimates from each to the production 2004 ACS estimates. Our objective was then to determine which of the two methods performed best overall at achieving the goals laid out including not introducing any unintended consequences with regards to the estimates and variances.

The most important goal was the equalization of householders and households. For the 2004 production data, the estimate of householders was 3.5 million greater than the estimate of households. For both alternative estimates (with and without subfamilies), that difference was reduced to approximately 70,000 nationally with the new methodology. This difference was also investigated at other levels of geography and the improvement was consistently demonstrated at all levels of county and above.

With our highest priority (and easiest to check) goal achieved, the remaining results examine the consistency between

household type and relationship and our review of other estimates including the household and housing unit estimates.

Household Type and Relationship

There were three goals to check for consistency between household type and relationship: spouses and married-couple households, unmarried partners and unmarried-partner households, and spouses in sub-families and married-couple subfamilies. Since these were all directly addressed in the raking methodology, the results here are a check on how compatible these goals were within the raking matrix. One primary objective in evaluating the results of these tables was to develop a recommendation on whether to include the equalization of married-couple subfamilies in the methodology or not based on the review of the estimates.

The goal of reducing the difference between the estimates of married-couple households and spouses had the highest priority because it is the largest group of the three. Table 1 shows the ACS 2004 production estimate, alternative estimate including the subfamily methodology, and the alternative estimate which does not include the subfamily methodology. The results in the table show that the alternative estimate reduces the difference between the estimate of spouses and married-couple households from 3.2% to 0.6% and 0.3% for the methods with and without married-couple subfamilies respectively. For context, the difference for the Census 2000 sample data is also shown. Both of the alternative estimates bring the difference between the number of married-couple households and the number of spouses to within sampling error unlike the ACS 2004 production which was highly significantly different. For comparing the two methods, we can disregard the sampling error on the difference between the estimate of spouses and married-couple households since the difference is due entirely to the person weighting methodology and not the data itself where the unweighted difference is zero. Thus, this data shows that the alternative estimate without subfamilies was more effective at reducing this difference than the method that included married-couple subfamilies and is more effective at the national level than the Census 2000 sample data which used a more complex raking methodology.

Table 1: Estimates of Married-Couple Households and Spouses with 90% Margin of Error

Type of Household and person-level estimate	ACS 2004 Production	Alternative Estimate with subfamilies	Alternative Estimate without subfamilies	Census 2000 Sample
Married-couple households				
Number of households	55,223,574	55,313,723	55,132,899	55,458,451
	(±290,689)	(±306,626)	(±286,942)	(±17,035)
Number of spouses	57,012,791	55,645,486	55,273,347	55,731,406
	(±180,015)	(±236,542)	(±237,369)	(±22,199)
Difference (Spouses-Households)				
Number	1,789,217	331,763	140,448	272,955
	(±341,914)	(±387,262)	(±372,397)	(±27,982)
Percent (Diff/households)	3.24	0.60	0.25	0.49
	(±0.62)	(±0.70)	(±0.68)	(±0.05)

The goal of reducing the difference between the estimates of unmarried-partner households and unmarried partners had a similar priority to the married-couple household goal although it is a much smaller group. Table 2 shows the equivalent results for the unmarried partners as the previous table did for spouses. While the improvements in the percent difference is not as large as for spouses in married-couple households, the difference from production is reduced from 3.3% to 1.7% and 1.3% for the methods with and without subfamilies in the methodology respectively. Like the married-couple households, the differences for each alternative method is now smaller than the sampling error associated with them. Again, the method which does not include the equalization for married-couple subfamilies performs the best.

The goal of reducing the difference between the estimates of married-couple subfamilies and spouses in subfamilies (a one-to-two relationship) had a lower priority than the other two household type and relationship equalization goals. This was because of the relatively small size of the estimate of married-couple subfamilies, the lower prominence of the two estimates (they are not shown in the main data profiles, for example), and because the ACS does not directly

Table 2: Estimates of Unmarried-Partner Households and Unmarried Partners with 90% Margin of Error

Type of Household and person-level estimate	ACS 2004 Production	Alternative Estimate with subfamilies	Alternative Estimate without subfamilies	Census 2000 Sample
Unmarried-partner households				
Number of households	5,840,833	5,663,939	5,637,879	5,230,703
	(±80,009)	(±81,432)	(±80,604)	(±7,404)
Number of partners	6,035,051	5,758,810	5,713,906	5,318,070
	(±83,677)	(±78,340)	(±77,826)	(±5,978)
Difference (Partners-Households)				
Number	194,218	94,871	76,027	87,367
	(±115,773)	(±112,997)	(±112,044)	(±9,516)
Percent (Diff/households)	3.33	1.68	1.35	1.67
	(±1.98)	(±2.00)	(±1.99)	(±0.18)

Table 3: Estimates of Married-Couple Subfamilies and Spouses in Subfamilies with 90% Margin of Error

Type of Household and person-level estimate	ACS 2004 Production	Alternative Estimate with subfamilies	Alternative Estimate without subfamilies	Census 2000 Sample
Married-couple subfamilies				
Number of subfamilies	891,721	882,490	883,028	940,784
	(±25,481)	(±26,055)	(±25,955)	(±3,242)
Number of spouses	1,907,217	2,045,164	2,348,755	1,868,518
	(±56,736)	(±60,860)	(±73,178)	(±4,770)
Ratio spouses/subfamilies	2.14	2.32	2.66	1.99
	(±0.09)	(±0.10)	(±0.11)	(±0.01)

capture the subfamily information (it is constructed through an edit). For these reasons it was viewed as a worthy yet expendable goal if it had an adverse or suboptimal impact on the other goals. Table 3 shows the relevant data for the consistency between the estimates of married-couple subfamilies and spouses in married-couple subfamilies. The logical ratio of spouses to married-couple subfamilies is exactly two and the table shows that both alternative methods actually increase the ratio away from 2.0 and away from the 2004 production estimate of 2.14. All three estimates are statistically significant from 2.0.

While the method that includes subfamilies does perform better than the method without subfamilies (2.32 versus 2.66), the subject-matter experts agreed that the better performance for subfamilies was not great enough to offset the gains that the method without subfamilies had for married-couple and unmarried-partner households. Thus, based on the estimates the subject-matter experts recommended the method which did not include subfamilies over the method which included subfamilies.

Other Population Characteristics

We also reviewed other key population characteristics across 34 person-based tables. Our objective was to note the number of significant differences between the production and the two alternative estimates, check for demographic consistency, and to compare the variance properties.

The number of significant differences between production and the two alternative estimates were principally concentrated in national-level estimates for relationship and marital status. At the state level, only approximately two percent of the changes in the estimates were statistically significant and at the county level only about 0.4 percent of the changes in the estimates were significant.

A non-parametric Friedman test was used to compare the variance properties for groups of similar characteristics between the two alternative methodologies and each alternative methodology as compared to the production variance estimates. This test showed no significant differences between the variance estimates for the two alternative methodologies. For the population characteristics, the only group that showed significant differences between the production and alternative methods was for the relationship and marital status tables.

The analysis based on all three criteria led to the same conclusion that apart from the household-type and relationship tables, both methods could be supported equally. Given the findings detailed above, this solidified the recommendation of the method which did not include subfamilies over the method which did. Further, the population subject-matter experts were able to fully recommend the implementation of the new weighting methodology.

Housing Units and Household Characteristics

Given the changes in the estimation procedure for households and the removal of the principal person adjustment from the housing unit weighting, much of our attention was on the housing unit and household characteristics tables. These tables made up 279 of the 313 tables studied.

Like the other population characteristics, the housing unit and household characteristic tables were checked to note the number of statistical differences and the variance properties for each of the alternative methodologies as compared to the 2004 production. In general, both analyses supported the two methods equally, however, both analyses noted problems in a subset of the housing unit tables which had a large number of significant differences and a significant increase in the variances. Further investigation revealed that while the new housing unit weight produced acceptable results for estimates of pure housing unit characteristics, those tables which crossed housing characteristics by person or householder characteristics showed a number of differences which were deemed unacceptable by the housing unit and household subject-matter experts. This difference had to be fixed in order to get their approval of the new methodology regardless of the benefits demonstrated for the person tables.

Table 4 shows the differences between estimates of households in the production and the alternative estimate. These estimates can be found both in a table of total occupied housing units by race of householder (ACS Table B25006) and in the race-iterated versions of the basic tenure table (ACS Table B25003A–B25003I). Because of this, these estimates are tabulated using the housing unit weight rather than the householder weight. The table shows the impact of having removed the principal person adjustment from the housing unit weight. Whites and American Indian or Alaskan Native (AIAN) show the greatest gains which are two groups that the ACS historically has had the strongest coverage relative to the population controls. The other groups all have substantial drops in their estimates.

The results of composite tables like these led to the exploration of other possible changes to the methodology specifically for housing unit tables which are crossed by household or householder characteristics.

Composite Housing Unit Tables and Vacancy

The first proposal for the composite housing unit tables was to use the person weight of the householder (the same weight used for the household tables) to tabulate all composite housing unit tables. To test this, the methodology for the alternative estimates without the use of subfamilies was applied using the 2005 ACS data and the 153 housing unit tables were tabulated using two sets of weights: one using the housing unit weight and one using the householder weight (for occupied housing units only). The estimates and variances from each run were compared to the 2005 production values. The results confirmed that the weight of the householder produced estimates for the composite tables that were more consistent with the 2005 production both with respect to the estimates and the variances. However, this introduced the possibility that tables like tenure could be tabulated using one weight for the normal tabulation of occupied/owned/rented housing units and a different weight for the same tabulation when crossed by the race of householder. This was recognized as simply transferring the original consistency problem of householders and households to a new area.

In reviewing all the data, however, it was found that using the weight of the householder also produced both acceptable estimates for the pure housing unit tables (those not crossed by any household or householder characteristics) while also having lower variances than using the alternative housing unit weight. After carefully reviewing all the housing

Table 4: Households by Race of Householder with 90% Margin of Error

Households by Race of Householder	ACS 2004 Production Estimate	Alternative estimate without Subfamilies	Percent Change from 2004 Production
All Households	109,902,090	109,569,818	-0.30%
	(±238,125)	(±271,588)	(±0.33%)
White	86,665,334	87,319,418	0.75%
	(±227,636)	(±364,930)	(±0.50%)
Black	12,963,975	12,293,076	-5.18%
	(±56,710)	(±216,644)	(±1.73%)
AIAN	770,686	808,937	4.96%
	(±21,505)	(±94,425)	(±12.57%)
Asian	3,966,989	3,800,053	-4.21%
	(±34,570)	(±68,807)	(±1.94%)
NHPI	116,315	111,589	-4.06%
	(±8,302)	(±11,457)	(±12.17%)
Some other race	4,053,538	3,901,153	-3.76%
	(±64,684)	(±101,175)	(±2.96%)
Two or more races	1,365,253	1,335,592	-2.17%
	(±39,758)	(±37,939)	(±4.03%)

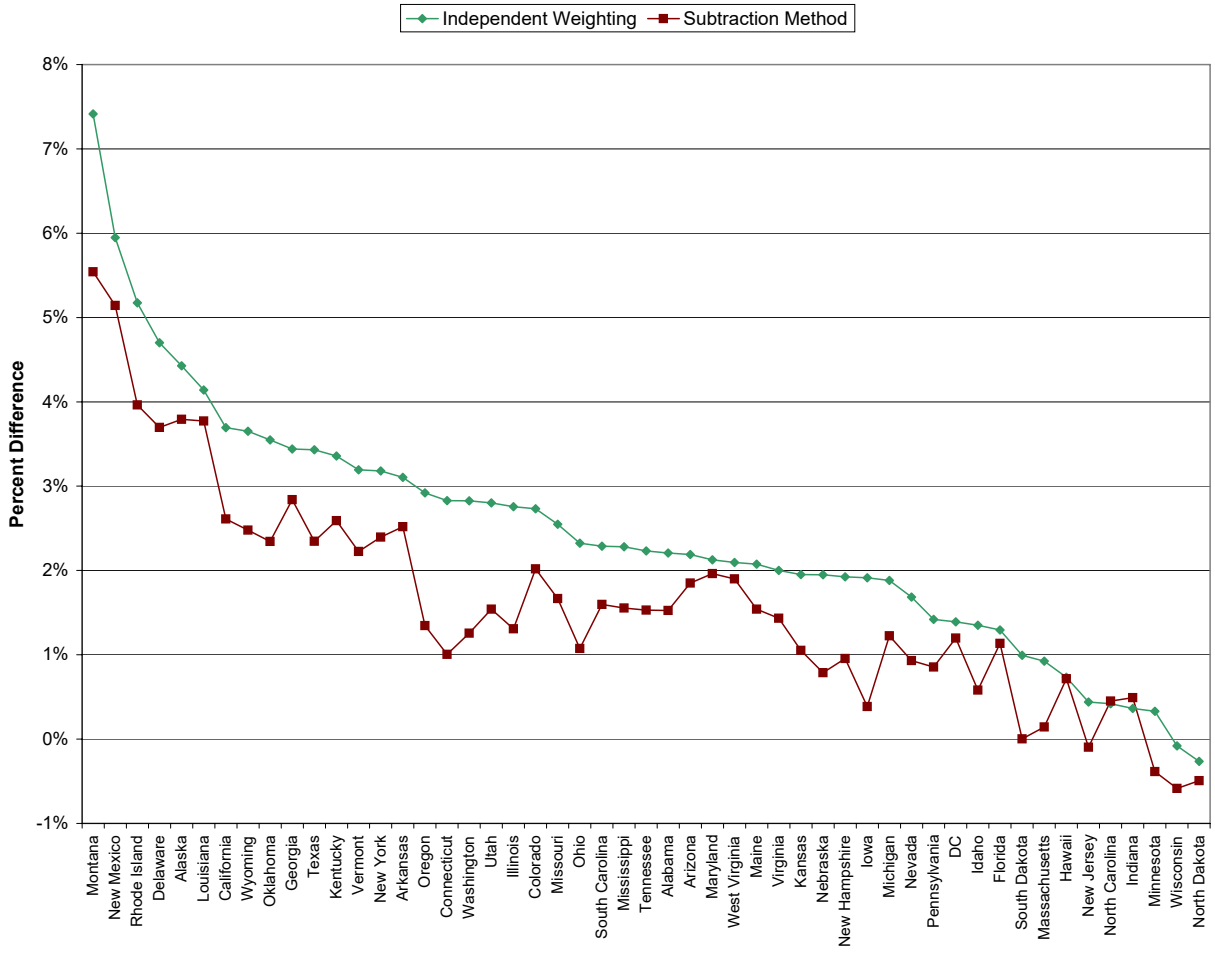
unit tables, the subject-matter experts recommended that the weight of the householder be used to tabulate all housing unit estimates for occupied housing units. This would ensure that the same weight is used to tabulate occupied housing units, households, and householders completely eliminating this data inconsistency.

The simplest way to accomplish this is to simply set the housing unit weight for all occupied units to be equal to the person weight of the householder, much in the way that was done for the principal person. The difference now is that the total person weight of the householders after the raking should be much closer to the original total housing unit weight for occupied housing units so the vacancy rate would not be impacted to the degree that it was with the principal person adjustment. This left the question of how to estimate the vacant housing units. In the methodology prior to 2006, a final ratio adjustment to the independent housing unit controls had been done to assure that the estimate of total housing units would be controlled. To use that method now would break consistency between the weight of the householder and the weight of the occupied housing unit. This left two choices: to back into the estimate of vacant housing units by subtraction of the estimate of occupied housing units from the total housing unit control or to leave the weight of the vacant housing units unchanged allowing the estimate of total housing units to float.

The latter choice would change what had been an objective of the ACS, that is, produce consistent estimates of total housing units with the Population Estimates Program (PEP), the Census Bureau's official source of total housing unit estimates. After internal discussions within the Census Bureau, it was agreed that since the raking would assure that the ACS and PEP housing unit estimates would be very close the decision should be made based on the quality of the vacant housing unit estimates produced by each method.

Figure 1 compares the the estimate of vacant housing units compared with ACS 2005 production by state. The graph is ordered from largest to smallest effect for the method which keeps the vacant housing units independent of the person weighting. Note that the largest difference using the 2005 data is for Montana with an increase of 7.5% in the number of vacant housing units to a slight decline for North Dakota. The highest percent differences are found in small states with the exception of California and the two lines generally follow each other. Overall, the independent vacant method causes a 2.3% increase in the estimate of vacant housing units nationally and the subtraction vacant method causes a 1.6% increase. Since the view of the subject-matter experts was that the ACS was underestimating the number of vacant housing units, the relatively higher estimate of vacant units in the independent vacant method

Figure 1: Differences in Vacant Housing Unit Estimates Compared with ACS 2005 Production, By State



led to their recommendation of that method over the subtraction vacant method.

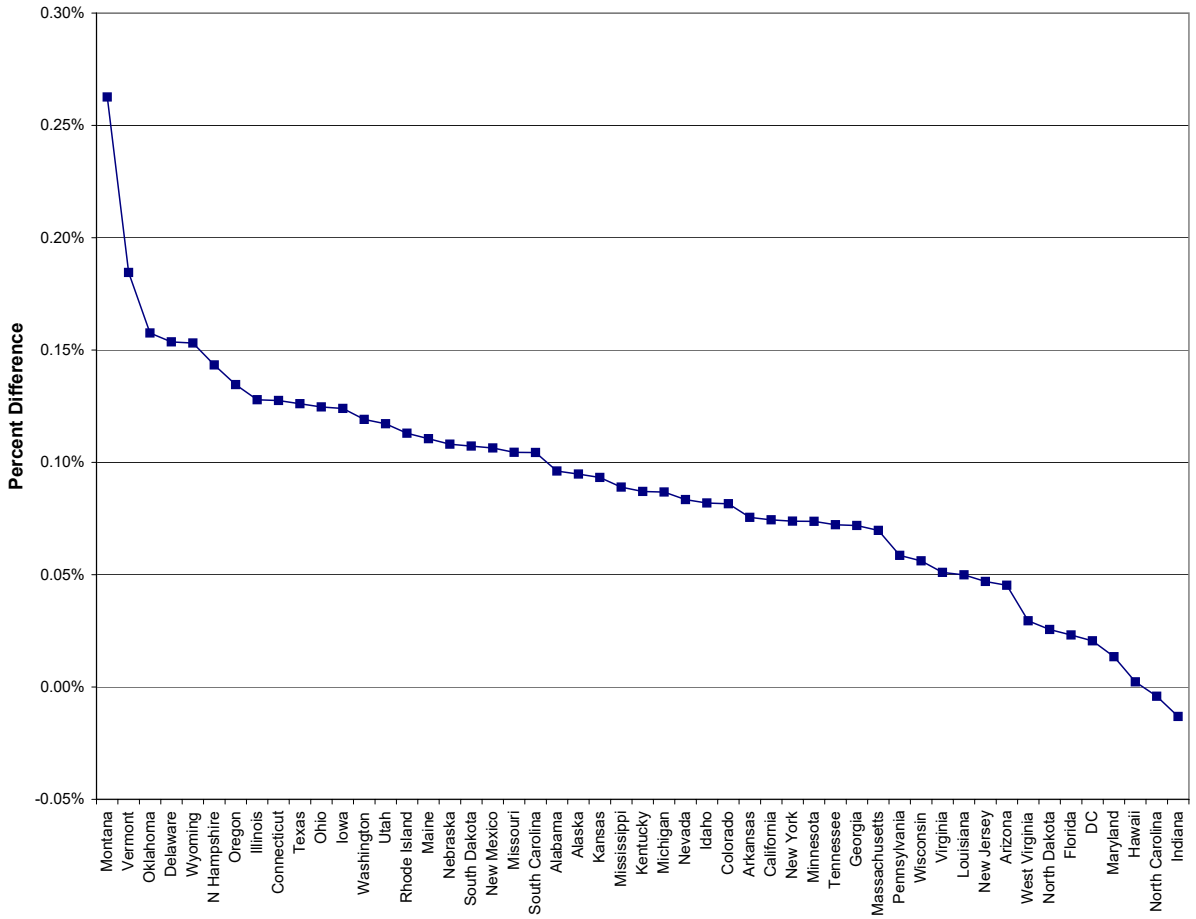
Since the independent method would cause the estimate of total housing units to no longer agree with the controls, this estimate was also investigated. Figure 2 shows the differences in total housing units from the controls by state. Again the states are ordered from largest positive effect to the lowest negative effect. Most states show a small increase in the estimate of total housing units with just a few small states showing an increase of 0.15% or more. Nationally, the difference is approximately 0.08% or about 97,000 housing units. This data acted to confirm that the small trade-off in the difference between the estimate of total housing units and the controls was worth the benefit in the improved estimates of vacant units. With these changes, the housing and household subject-matter experts were able to fully recommend the implementation of the new weighting methodology.

Implications and Conclusions

With the final methodology in place, all of the goals were met with the exception of the married-couple subfamilies. In the end the improvements gained toward the other goals without including the portion of the methodology that addressed the married-couple subfamilies outweighed the gains by including it. This new methodology was implemented in the 2006 ACS weighting released in the summer of 2007.

To summarize the results along the lines of the goals outlined in the Research Goals section:

Figure 2: Differences in Total Housing Unit Estimates Compared with ACS 2005 Production, By State



1. The estimate of householders will be exactly equal to the estimate of households (which is exactly equal to the estimate of occupied housing units).
2. The difference between the estimates of spouses and married-couple households was cut from 3.6% to 0.3% based on 2004 data.
3. The difference between the estimates of unmarried partners and unmarried-parter households was cut from 3.3% to 1.3% based on 2004 data.
4. We expect that the difference between the estimate of married-couple subfamilies and spouses in subfamilies to be no larger than the empirical results shown with the 2004 data.
5. The overall estimate of vacant housing units should rise approximately 2.3% nationally based on 2005 data.
6. There should not be any systematic significant changes to other tables neither in their estimates nor in their variances with the exception of total housing units which will no longer be exactly equal to the controls at the weighting area level or higher (including states). There will be a small margin of error associated with this estimate to reflect the fact that the final estimate of total housing units is no longer controlled.

Based on these results for the 2004 and 2005 data, we expected that the impacts on the 2006 data to be similar. Given that the impacts above are substantial, the plans for the implementation of this methodology were shared through various user groups and a user note was drafted to accompany the 2006 data release. Overall, we received very positive feedback from users concerning our plans.

We also examined how efficiently the raking matrix was performing for the 2006 weighting. Using the adopted methodology, we used 19 iterations everywhere based on identifying the point at which nationally the data inconsistencies cease to improve and actually begin to worsen. Our plan is to automate the process of determining how many iterations to use by adding a convergence criterion possibly at the weighting area level rather than the national level. A topic for future research is to see if we can collapse the raking down to two dimensions in place of three. Our initial work shows that the raking could converge both more quickly and closer to the control totals using two dimensions in place of three. Before implementing such a change, however, we would need to repeat our analysis that changing the number of dimensions would not cause an adverse impact on either the estimates or the variances. In particular, we are concerned that the variances could increase using this method.

References

Alexander, C. & Roebuck, M. (1986). Comparison of Alternative Methods for Household Estimation, Proceedings of the 1986 American Statistical Association, American Statistical Association, Alexandria, VA.

Navarro, A., Albright, K. & Asiala, M. (2004). An Alternative to the Principal Person Method for Weighting in the American Community Survey, Proceedings of the 2004 Joint Statistical Meetings, American Statistical Association, Alexandria, VA.

Navarro, A., Griffin, R., & Bates, L. (1990). Family Estimates Empirical Study, Proceedings of the 1990 Meetings of the American Statistical Association, American Statistical Association, Alexandria, VA.

US Census Bureau (2006). Design and Methodology: American Community Survey, US Government Printing Office, Washington DC.