

Household Estimates Conundrum: Effort to Develop More Consistent Household Estimates Across Surveys

Arthur R Cresce, Yang Cheng, and Christopher Grieves*

U.S. Census Bureau
Washington, D.C. 20233

Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference

Introduction

One of the most basic demographic concepts is that of households. If one were to look at the technical documentation for the decennial census and our current surveys (see attachment A), one would find that we consistently define households as including all the people who occupy a housing unit. Thus, by definition, the number of households (householders) should be equal to the number of occupied housing units. However, despite the consistency of definition of households (occupied housing units), the Census Bureau has had to deal with the issue of different estimates of households in the United States among its various data collection efforts. Table 1 shows these differences among 4 surveys: the Current Population Survey/Annual Social and Economic Survey (CPS/ASEC), the Housing Vacancy Survey (HVS - part of the CPS), the American Community Survey (ACS), and the American Housing Survey (AHS).

Table 1 Estimate of Households and Occupied Housing Units in the Current Population Survey, the Housing Vacancy Survey, and the American Housing Survey, and the American Community Survey

(Numbers in thousands)

	2003	2004	2005	2006	2007	2008	2009	2010[1]	2011	2012
CPS Basic	112,312	113,291	114,475	115,942	117,474	118,033	118,518	118,636	119,571	121,804
CPS/ASEC [2]	111,278	112,000	113,343	114,384	116,011	116,783	117,181	117,538	118,682	121,084
CPS/HVS	106,111	107,606	109,484	110,693	111,258	111,672	112,295	112,899	113,533	114,513
ACS	108,420	109,902	111,091	111,617	112,378	113,101	113,616	114,567	114,992	115,970
AHS	105,842	NA	108,871	NA	110,692	NA	111,806	NA	114,907	NA

[1] Further complicating the situation is the fact that the 2010 Census provided yet another total for households that was different from the other estimates show below:

2010 Census	116,716
CPS Basic	118,636
CPS/ASEC	117,538
CPS/HVS	112,899
ACS	114,567

We have attempted to explain the differences between the 2010 Census and the 2010 1-year ACS, Griffin (2011) and Cresce (2012) provided a preliminary analysis of possible sources of differences, including comparison of field procedures; evaluations of 2010 Census methods; assessment of ACS methods, frame, sampling and estimation; analysis of results from a match of 2010 Census and 2010 1-year ACS records; and analysis of aggregate data from both sources. It is clear at this stage that there are factors associated with each data source that can help explain at least a portion of the differences noted.

[2] Although the Housing Vacancy Survey is part of the Current Population Survey and is conducted by the same field representatives, the results for vacant and occupied housing units from the Housing Vacancy Survey are based on independent housing unit controls while the results for the occupied housing units (households) in the CPS/ASEC are based on independent population controls. The CPS/HVS estimates are based on vintage 2011 housing unit controls provided by the Population Division.

*The views expressed in this paper are solely attributable to the authors and do not necessarily reflect the position of the United States Census Bureau.

It is clear from an initial reading of Table 1 that, with a few exceptions, none of these sources of data on households agree. Across all years, there is a consistent pattern with the CPS/ASEC providing the highest estimate, the ACS providing the second highest estimate, and the CPS/HVS providing the lowest estimate. In odd-numbered years when the AHS is conducted, the AHS estimate is not different from the CPS/HVS in 2003 and 2005 and the estimate for the 2011 is not statistically different from the ACS estimate for that year. It is not surprising, then, that data users are confused and unsure what estimate to use for their programs/analyses.

If the concept underlying these estimates is the same across the three surveys, why should there be such large differences? The answer to this question is largely due to: 1) differences in estimation methods, 2) differences in sample design, and 3) differences in operational procedures. The focus for this paper will be on: 1) understanding the reasons for the differences among the current surveys, including the American Community Survey and 2) developing a methodology for calculating consistent household estimates among the various CPS-based surveys ("basic" CPS, CPS/ASEC and the CPS/HVS). The decennial census, as noted in Griffin (2011) and Cresce (2012), poses unique challenges in trying to understand these differences but will not be the focus of this paper. We expect that the same (or similar) approach for obtaining consistent household estimates for the CPS would be employed at some point for the American Housing Survey and the Survey of Income and Program Participation (SIPP), with some possible revisiting of the ACS methodology to make sure the approaches are as similar as possible. Furthermore, although we state that the differences in household estimates are a function of both estimation issues, design issues, and operational issues, we believe that most of the differences are due to estimation and will only mention possible design and operational issues while focusing primarily on estimation issues.

Differences in Estimation

In the early 2000's, an interdivisional team of statisticians at the Census Bureau examined the use of housing unit controls versus population controls as it applied to Housing Vacancy Survey (see Love, 2003). They concluded that the use of the weight adjustment for within household undercoverage when using population controls by age, sex and race tended to be higher than the adjustment to the weights for housing unit coverage when using housing unit controls, which focus on coverage of housing units (including whole households). Thus, when population control-based weights are applied to characteristics such as household relationship, the estimate of householders (family plus nonfamily) will almost always be higher than the corresponding housing unit control-based weights that are applied to obtain the estimate of occupied housing units.

This apparent discrepancy had been an issue when results for estimates of occupied housing units were compared between: 1) the Housing Vacancy Survey (before 2002), which is part of the Current Population Survey and had been based on population control-based weights and 2) the American Housing Survey¹, which was based only on housing unit control-based weights, and the New York City Housing and Vacancy Survey², which was based on a combination of housing unit control- and population control-based weights. Ultimately, discussions about this difference led to the Housing Vacancy Survey applying housing unit control-based weighting that not only resulted in consistent estimates between the HVS and the AHS, but also resulted in a precipitous drop in the estimate total housing units and occupied units in 2003 from the old method to the new method (See Table 2 below).³

¹ Since the American Housing Survey primarily focuses on housing characteristics, no population weights are used and any population characteristics shown are those of the householder or the household. In addition, we should note that AHS estimates are controlled to selected CPS/HVS estimates for householders and for vacancy status.

² In fact, the New York City Housing Vacancy also had this problem (of inconsistent occupied units vs. households) until the sponsors insisted that the estimates be brought into agreement and the Statistical Methods Division devised a method to achieve this goal.

³ The CPS/HVS estimation methodology essentially takes the base population weight that has been adjusted for differential undercoverage using population controls and controls those weights to add up to the independent housing controls and apportions the remaining weight to other household members. See "Memorandum for Chester E. Bowie from Alan R. Tupek, 'Weighting Specifications for the Housing Vacancy Survey (HVS2000-1)', June 1, 2004. This approach appears to be similar to the approach used for ACS estimation after 2000, but this observation needs to be verified.

Table 2. Comparison of HVS Estimates Based on Population-Based vs Housing-Based Weights			
	2003 - Housing Based*	2002 - Housing Based	2002 – Pop Based
Total Housing	120,834	119,297	123,318
Occupied	105,560	104,965	108,539
Vacant	15,274	14,332	14,779

*The estimates for 2003 in this table are based on 2001-vintage housing unit controls. Thus, the estimate of households (occupied housing units) in this table will differ from that in Table 1.

The American Community Survey (ACS) also dealt with this issue. Since the ACS produced both population and housing estimates, it had to address some troublesome internal inconsistencies between estimates of occupied housing units and households. This issue was partially addressed by using housing unit controlled weights for housing unit tables and tables where households were the universe. However, from the inception of the ACS through 2005 there still remained the problem that estimates of households obtained by adding family and nonfamily householders from the relationship question (based on population controlled weights) were still much higher (as expected) than the estimates for occupied housing units and households. For example, in the 2005 ACS, the estimate for occupied housing units was 111.1 million while the corresponding estimate for householders was 114.8 million. This problem was addressed in 2006 with the application of a new estimation methodology (for a full explanation, see:

http://www.census.gov/acs/www/Downloads/survey_methodology/acs_design_methodology_ch11.pdf) that essentially guaranteed consistency between estimates of occupied housing units and householders in housing unit tables and estimates of householders in population tables by creating a householder weight that could be used both for housing and population tabulations. One of the distinctive features of this new methodology was the fact that estimates for both the vacant and occupied housing units were to be obtained directly from the survey, meaning: 1) that their sum would not necessarily add to the housing control total and 2) the housing unit total would now have sampling error attached to it.

However, despite the progress made in harmonizing occupied housing unit/householder estimates for internal consistency within the ACS and the New York City Housing and Vacancy Survey, we must address the following questions:

- 1 Should we apply a similar type of methodology to the base CPS, the CPS/ASEC, and the CPS/HVS to ensure reasonably consistent household estimates and what are the consequences of implementing such a methodology for some of the key statistics produced by these surveys?
- 2 Will the resulting methodology used in 1 above be comparable with those employed in the ACS and the New York City and Housing Vacancy Survey?
- 3 If they are different, can that partially account for the differences among them and can they all be reconciled into one methodology that can be applied for all these surveys?

This paper will not address the second and third questions, which, however, ultimately must be answered after we address the first question regarding the potential application of the same housing-unit based estimate of households (occupied housing units) across the base CPS, the CPS/ASEC, and the CPS/HVS.

Differences in Sampling Frame

Related to the issue of estimation is the fact that the ACS, CPS and the AHS are based on somewhat different sampling frames. The ACS is solely based on the master address file (MAF), which was significantly updated using results of the 2010 census. The current CPS sample is based on a sampling frame originally designed using addresses from the 2000 Census and updated during the decade for new construction. The new CPS sample design, which will be implemented starting in April 2014, will be based on the MAF. The AHS is based on a sample that was originally designed from an address file resulting from the 1980 census with updates for new construction ever since. Starting in 2015, the AHS will also be based on an entirely new sample drawn from the MAF. It is not clear what impact differing sample frames for each of these surveys has had on differences among these surveys for

characteristics or concepts that presumably the same for all three surveys⁴. However, to the extent that differing sample frames might explain differences observed in estimates for households (occupied housing units), this potential source of differences could be substantially mitigated.

Differences in Operational Procedures

While we believe that the impact of estimation differences probably explains the largest portion of the differences among the surveys in estimates of households (occupied housing units), operational differences also help explain differences in household estimates. Although we have noted above that the ACS has been able to achieve internal consistency among the estimates of occupied housing units, households and the sum of family and non-family householders, this survey regularly produces estimates of vacant units that are lower than those produced by the HVS and the AHS. Conversely, the estimate of occupied housing units (households) is almost always higher in the ACS than in the HVS or the AHS. This phenomenon is well known and was documented in a report by Schwartz (2009) comparing results of the American Housing Survey with those from the ACS.⁵ In this report, the gross vacancy rate for the ACS was 12.1 percent – 1.6 percentage points lower than the AHS gross vacancy rate of 13.7 percent. The tendency for the ACS to produce lower estimates of vacant units (and higher estimates of occupied units) than the HVS or the AHS can be attributed, in part, to the fact that in the ACS, vacant units are mostly identified in the personal visit stage of enumeration, which occurs up to three months after the initial mail out. This procedure, coupled with a “current residence” rule of 2 months, could result in classifying housing units that were vacant at the initial mail out being occupied by the time of the personal visit interview. This phenomenon is much more likely to occur in rental units than in owned units.⁶ Thus, the estimate of households (occupied housing units) from the ACS will almost always be higher for the ACS than for the AHS or the HVS. In addition, it is also possible that the ACS sampling and estimation procedures for vacant units may also contribute to these differences since vacant units are mostly included in the nonresponse universe, which is subsampled at a given rate and results are weighted at a much higher rate than results from the mail form.

There appears to be no really significant differences in operational procedures between the CPS/HVS and the AHS. In fact, they have in common something that is not done in the ACS (nor in the decennial census) in that they both determine for vacant units whether the unit is intended for year-round use. If the housing unit is intended for year-round use, the Field Representative (FR) then asks the vacancy status questions (that is, is the housing unit for sale, for rent, or some other reason). If the vacant housing unit is not intended for year-round use (even if it is capable of being occupied at any time of the year), the unit is automatically classified as “seasonal” and no attempt is made to determine if it is for sale or rent. ACS FRs and decennial census enumerators, on the other hand, do not ask this question and determine vacancy status for all vacant housing units. We need to determine if we should make the ACS and the decennial census conform to the HVS and the AHS or vice versa, even though we believe the differences do not affect differences in occupancy status.

One other possible source of differences that we need to explore is the possible use of different rules or procedures for determining occupancy status in current surveys. We know from discussions with FRs that determining occupancy status can be one of the most difficult tasks that FRs must accomplish, especially if they are having difficulty finding a respondent or a “knowledgeable source”. The FRs success in obtaining this information for a particular housing unit will help determine whether the unit is vacant or a refusal (Type A). The consequences of designating a housing unit as vacant or Type A are significant since Type A noninterviews are accounted for by a noninterview adjustment in the very early stages of the estimation process while an erroneous designation of a housing unit as being vacant results in a whole household miss with no adjustment for loss of coverage of the household. We need to review practices across surveys to ensure that consistent rules are being applied across surveys.

⁴ We have some evidence from Hefter and Anderson (2012) of the possible impact of differing frames between the 2010 American Community Survey 1-year data and the 2010 census. However, this paper did not assess how that impact might explain differences between the 2010 Census and the 2010 1-year ACS in gross vacancy rates.

⁵ Ironically, the gross vacancy rate from the 2005 New York City Housing Vacancy Survey (NYCHVS) was lower than the 2005 ACS estimate for New York City. However, the authors note that there are various reasons why the NYCHVS estimate would be lower than the ACS estimate (see Callis and Fronczek, 2008).

⁶ For example, in 2007 the ACS rental vacancy rate was 7.9 percent, compared with 9.8 percent for the AHS, while the ACS homeowner vacancy rate was 2.5 percent, compared with 2.6 percent for the AHS (Schwartz, 2008, Table 2a).

We will continue to research these issues to see what impact they might have on differences among the surveys. The remainder of this paper, however, will focus on efforts to address estimation methodology for households (occupied housing units) in the CPS.

Developing a Housing-Unit Based Methodology for Estimating Households in the CPS

The intentions of our research is to try a modification to the current base CPS weighting scheme by introducing house unit controls. We hope that this will produce consistent estimates of the number of occupied households and their characteristics among CPS related surveys. At the same time, we will monitor the impact of these methodologies on a wide range of characteristics, most particularly labor force, income and poverty statistics.

Beyond just the immediate effect on labor force statistics, we have to account for the fact that other surveys are stakeholders in the results of base CPS. For example, three immediate stakeholders would be: 1) the CPS/HVS, which uses the Base CPS “Family Weight” in its estimation procedure, 2) the CPS supplements (especially the CPS/ASEC), and 3) the AHS which, in turn, controls to proportions of demographics created using CPS/HVS and base CPS data over the same time period during which AHS field enumeration is being conducted. These are all things we must take care to acknowledge while conducting our research and take care not to introduce any bias into our estimates.

One way we may introduce the total housing unit control to estimate occupied housing units is as follows. After the State Coverage Adjustment in the CPS, we create an estimate of total housing units from the survey using the householder’s weight and the base weights for the vacant homes. Then using the ratio of the housing unit control over the estimate of total housing units we have created, we can ratio adjust the weights of all the householders. We can obtain a new estimate as follows

$$\widehat{O}^* = \frac{\widehat{O}}{\widehat{O} + \widehat{V}} \times HUC = \widehat{O} \times f_{hh} = \sum_{i \in S} w_{SC\ i} I_i \times f_{hh}$$

$$I_i = \begin{cases} 1 & \text{if } i \text{ is a householder} \\ 0 & \text{if } i \text{ is not a householder} \end{cases}$$

Where \widehat{O}^* is our new estimate of occupied households, \widehat{O}, \widehat{V} are estimates of occupied and vacant housing units, HUC is the total housing unit control, and $w_{SC\ i}$ is the weight of the reference person in the above summation. S is the selected sample from the non-institutionalized labor force.

We then create an estimate of the Non-Householders, or “Residual” population by

$$\widehat{Res}^* = Pop - \widehat{O}^* = \widehat{Res} \times \frac{Pop - \widehat{O}^*}{\widehat{Res}} = \widehat{Res} \times f_{res} = \sum_{i \in S} w_{SC\ i} I_i \times f_{res}$$

Where

$$\widehat{Res} = \sum_{i \in S} w_{SC\ i} I_i$$

and

$$I_i = \begin{cases} 1 & \text{if } i \text{ is not a householder} \\ 0 & \text{if } i \text{ is a householder} \end{cases}$$

\widehat{Res}^* is our new estimate of the non-householder population, \widehat{Res} is the current estimate of the non-household population, Pop is the total population control for the civilian non-institutional labor force, and $w_{SC\ i}$ are the state coverage weights summed over non-householders.

However, in order for these estimates to hold through second stage, we would need to constrain our householder and non-householder population estimates to \widehat{O}^* and \widehat{Res}^* , respectively. To accomplish this, we include a new dimension in the second stage raking matrix with margins containing \widehat{O}^* and \widehat{Res}^* for the householders and non-householders respectively. All other margins remain as they currently are.

The above method needs some refinement for the CPS. Some other considerations are:

1. What level of geography should we apply the adjustment at? National, Regional, State?
2. CPS does weighting by month-in-sample pairs (1,5), (2,6), (3,7), (4,8). Should we give each pair an equal proportion of our developed control?
3. Our Occupied Housing Unit “control” in the Second Stage Raking procedure has variance. What affect does this variance have on point estimates and variance estimates?
4. The above method gives estimates closer to those from other surveys, especially the CPS/HVS. However, the main concern is inference on the population’s labor force characteristics and coverage of the different demographics. As base CPS is only concerned about the civilian non-institutionalized labor force, does this give us correct inference of the population and make sense? Are there actual coverage issues with householders vs non-householders?

We performed the above method by performing the above adjustment at the state level, including the District of Columbia, for July 2012. We gave each month-in-sample pair an equal proportion of the control for occupied housing units. The table below provides the results of applying this approach.

Table 3. July 2012 Estimates of Occupied Housing Units New Method\Current Weighting

Housing and Population Controls	Population Controls Only
113,778,859	121,854,557

We see that performing the above procedure gave an estimate of occupied housing units (households) that was considerably lower than the estimate obtained from using only population controls but was slightly lower estimate of the number of households than those from other surveys. However, more analysis needs to be done to address its impact on coverage and employment characteristics. We also need to address the level of aggregation.

We also note that this is not the only approach one can take. Creating a separate weight for households and persons can make more sense in the context of doing inference on characteristics of households. We plan to do future research on this topic.

Recommendations and Conclusion

This discussion and the decisions we make are very important for the estimate of households and vacant housing units. We know that the housing industry is very interested in the number of households and in the growth of households for the purposes of residential construction. In addition, there are certainly many other uses for accurate data on household, both by private industry and by federal, state and local governments. The fact remains, however, that somehow our estimates of households (occupied housing units) and, by definition, our estimates of vacant housing units are not in sync.

We recommend that the following actions be taken:

1. Develop alternative estimation methodologies for the base CPS and determine the impact of these methodologies: a) on various key statistics produced by the base CPS and b) on the resulting estimate of households (occupied housing units).
2. Take the preferred methodology developed in 1 above, implement that methodology for the CPS supplements and the CPS/HVS and assess the impact of this methodology: a) on various key statistics produced by these CPS surveys and b) on the resulting estimate of households (occupied housing units)
3. Review systematically the procedures for determining the occupancy status of housing units across current surveys and from one Regional Office (RO) to another. “Best practices” should be learned and applied across all surveys in each RO. We are already working to make sure that more detailed categories for the “Other Vacancy” status category are being applied across our housing surveys.

4. Determine whether the “year-round use” concept for housing units should be used not only for the HVS and the AHS but also for all our other surveys where FRs need to determine occupancy status or whether it should be dropped altogether.
5. Investigate differences between household estimates obtained from our newly developed household estimation methodology for the CPS with those from the ACS and those from the New York City Housing and Vacancy Survey (NYCHVS). If necessary, adjust the methodologies where necessary to ensure that all methodologies are employing the same (or similar) approach.
6. Apply this same methodology to the Survey of Income and Program Participation and assess the impact of using this methodology on key statistics produced by the survey.

Conclusion

We have an obligation to our data users to provide clear, unambiguous explanations and guidance about the estimates we provide on households (occupied housing units). In our role as the “leading source of quality data about the Nation’s people and economy,” we ultimately should be able to:

1. Develop a common methodology that produces reasonably consistent estimates of households (and, therefore, vacant housing units) across our current surveys,
2. Develop a common methodology that may still produce noticeable differences but still be able to explain differences and recommend a preferred estimate, or
3. Be able to explain why we cannot develop a common methodology and be able to explain the differences and provide guidance about the agency’s preferred estimate of households (occupied housing units) for the United States.

References

- Cresce, Arthur. R. 2012. Evaluation of Gross Vacancy Rates From the 2010 Census Versus Current Surveys: Early Findings from Comparisons with the 2010 Census and the 2010 ACS 1-Year Estimates. Paper presented at the January 2012 meeting of the Federal Committee on Statistical Methodology in Washington, DC. Available at: <http://www.census.gov/housing/files/FCSM%20paper.pdf>.
- Griffin, Deborah. 2011. Comparing 2010 American Community Survey 1-Year Estimates of Occupancy Status, Vacancy Status, and Household Size with the 2010 Census - Preliminary Results. Available at: http://www.census.gov/acs/www/Downloads/library/2011/2011_Griffin_03.pdf.
- Schwartz, Mary B. 2009. 2007 American Community Survey: A Comparison to Selected Housing and Financial Characteristic for the United States from the 2007 American Housing Survey. ACS evaluation report available at: http://www.census.gov/acs/www/Downloads/library/2009/2009_Schwartz_01.pdf.

Attachment A – Definitions of Households

Decennial Census

A household includes all of the people who occupy a housing unit. A housing unit is a house, an apartment, a mobile home, a group of rooms, or a single room occupied (or if vacant, intended for occupancy) as separate living quarters. Separate living quarters are those in which the occupants live separately from any other people in the building and that have direct access from the outside of the building or through a common hall. The occupants may be a single family, one person living alone, two or more families living together, or any other group of related or unrelated people who share living quarters.

ACS

A household includes all the people who occupy a housing unit. (People not living in households are classified as living in group quarters.) A housing unit is a house, an apartment, a mobile home, a group of rooms, or a single room that is occupied (or if vacant, is intended for occupancy) as separate living quarters. Separate living quarters are those in which the occupants live separately from any other people in the building and which have direct access from the outside of the building or through a common hall. The occupants may be a single family, one person living alone, two or more families living together, or any other group of related or unrelated people who share living arrangements.

CPS/ASEC

A household consists of all the people who occupy a housing unit. A house, an apartment or other group of rooms, or a single room, is regarded as a housing unit when it is occupied or intended for occupancy as separate living quarters; that is, when the occupants do not live and eat with any other persons in the structure and there is direct access from the outside or through a common hall.

A household includes the related family members and all the unrelated people, if any, such as lodgers, foster children, wards, or employees who share the housing unit. A person living alone in a housing unit, or a group of unrelated people sharing a housing unit such as partners or roomers, is also counted as a household. The count of households excludes group quarters. There are two major categories of households, "family" and "nonfamily". (See **definitions of Family household and Nonfamily household**).

HVS

Occupied Housing Units. A housing unit is occupied if a person or group of persons is living in it at the time of the interview or if the occupants are only temporarily absent, as for example, on vacation. The persons living in the unit must consider it their usual place of residence or have no usual place of residence elsewhere. The count of occupied housing units is the same as the count of households.