# Panel Analysis of Household Nonresponse and Person Coverage in the Current Population Survey

Edwin Robison[1], Christopher Grieves[2]
[1]US Bureau of Labor Statistics
[2]US Census Bureau

## Abstract

Nonresponse and coverage problems affect every household survey, but it can be difficult to determine the impact on estimates. The Current Population Survey is a monthly panel survey, and each panel is included in the survey 8 months. The panel aspect of the survey expands the possibilities for analyzing nonresponse and coverage. For example, panel data on relationships of household members to head-of-household can be examined to determine if relationship structures systematically change during the eight months. Another possibility is to link panel microdata across 8 months. The linked microdata can be used, for example, to measure differential response (and presumably nonignorable nonresponse) among households with different relationship structures. Results from analyses of panel data are presented and suggestions are made for further research.

**Key Words:** Nonignorable nonresponse, attrition, coverage

## 1. Introduction

The monthly Current Population Survey (CPS) is jointly sponsored by the US Bureau of Labor Statistics (BLS) and the US Census Bureau.[1] It is highly visible since it produces the national unemployment rate (UER), designated by the US Office of Management and Budget as a Primary Federal Economic Indicator (PFEI). The CPS is a key contributor to measures that are used to determine the distribution of some federal funds to states and localities. The survey has a long stable history and, due to the importance of the CPS, the agencies are very careful and conservative about making methodological changes.

Being conservative about change does not preclude change, even major changes. The current research on household response and person coverage is in its early stages. Some interesting results have been found, but no methodological changes are proposed. However, the current work is akin to research a decade ago on differential coverage by race and ethnicity that resulted in a restructuring of CPS weighting (Robison, 2003).

The two agencies always have a general interest in household response and person coverage. There is heightened interest due to these factors:

- Household nonresponse slowly increased from 7% in 1995 to 11% in 2014
- Panel bias has changed
- Household estimates made from the CPS seem too large compared to other surveys

The flow of this paper will be as follows.

[1] Views expressed are those of the authors and do not necessarily reflect the views or policies of the Bureau of Labor Statistics or the Census Bureau.

- Summary information on CPS sampling and labor force (person) estimation
- Some interesting research results on differential response (attrition in this case)
- The context of housing unit or household estimation (as opposed to person or labor force estimation)
- Possibilities for future research.

## 2. Summary of CPS Sampling, Weighting, and Estimation

The Current Population Survey is a monthly survey primarily aimed at collecting labor force information for the United States and for each individual state. The target population is the Civilian Noninstitutional Population (CNP), and labor force information is restricted to adults 16 years of age and older. What follows in this section is a summary of current CPS methodology; refer to Technical Paper 66 for details (Census, 2006).

### 2.1 CPS Sampling

A full monthly sample of about 77,000 "designated" housing units (HUs) is a representative probability sample of a frame of about 135,000,000 HUs. The sample can generally be described as a two-stage stratified probability sample – large areas called Primary Sampling Units (PSUs) are selected, and then housing units are selected from those PSUs. The most populous PSUs in each state are sampled with certainty and are called self-representing PSUs; the remaining PSUs are probability sampled and are called non-self-representing PSUs. In addition to national data reliability needs there are also state reliability needs, so samples in less populous states are denser than samples in more populous states

Only the ≈60,000 housing units occupied by at least one adult in the CNP are included in data collection. In 2012 and 2013 about 53,500 housing units responded monthly. Each month detailed information including labor force data is obtained for about 105,000 adults. CPS survey operations determine which housing units are occupied, and which ones are permanently or temporarily out of scope. Note that the CNP is more restrictive, or smaller, than the resident population used for other Census surveys such as the American Community Survey (ACS) and the Survey of Income and Program Participation (SIPP). Some housing units that would be considered occupied for those surveys would be considered out-of-scope housing units for the CPS (so CPS would not consider them to be occupied).

The CPS sample is divided into 8 panels. The panel structure is exploited for the analysis leading to the results presented in this paper.

- Any given panel is interviewed four consecutive months, is dropped for 8 months, and then is brought back for another 4 consecutive months of interviewing (a total of 8 months in the sample, denoted MIS1-MIS8).
- Every month has one panel that is in each of the 8 months in sample.
- Each month two panels are rotated out and two panels are rotated in, so 6 of the 8 panels are in common from one month to the next. (This overlap is used in composite estimation that will be briefly discussed.)

The full rotation chart is not given, but the following schematic shows what happens between adjacent months t and t+1. The panel in sample for the first time in month t is designated MIS1 and it continues into month t+1 and is in sample for the second time: MIS1→MIS2. For the other continuing panels: MIS2→MIS3, MIS3→MIS4, MIS5→MIS6, MIS6→MIS7 and MIS7→MIS8. The month t panel in MIS8 does not

continue, is permanently dropped, and is replaced in month t+1 by a new panel in MIS1. The month t panel in MIS4 is temporarily dropped for 8 months and in month t+1 is replaced by a returning panel in MIS5.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| month t | MIS1 ↓ | MIS2 ↓ | MIS3 ↓ | MIS4 | | MIS5 ↓ | MIS6 ↓ | MIS7 ↓ | MIS8 |
| month t+1 MIS1 | MIS2 | MIS3 | MIS4 | | MIS5 | MIS6 | MIS7 | MIS8 | |

## 2.2 CPS Weighting

There are several stages in CPS weighting. Initial weighting steps are housing-unit based and all persons in a HU will have the same weight.

- The base weight for a HU is the inverse of its probability of selection.
- There is a housing-unit-based nonresponse adjustment procedure. Weighting class ratio adjustment is used, where the adjustments are of the form (r + Type_A)/r. "Type_A" HUs are those that are determined by survey operations to be occupied but not responding. The weights of occupied HUs r in a cell are increased to cover all occupied HUs in the cell.

Subsequent weighting steps are person based, and different persons in a housing unit will have different weights. Four procedures use forms of ratio adjustment or benchmarking that force weighted CPS data to match population "control" figures provided by the Census Bureau (controls derived externally to CPS). A fifth ratio adjustment procedure, composite estimation, uses a combination of external and internal controls. In second-stage ratio adjustment and composite weighting the controls are used for each panel/MIS.

- First-stage ratio adjustment factors are applied only to sample data in non-self-representing Primary Sampling Units (NSR PSUs). In each state factors are computed by race to align the race characteristics of sampled NSR PSUs with the known race characteristics of all NSR PSUs. The procedure has only a modest impact on labor force estimates or on estimates of households.
- National coverage adjustment groups data into cells defined by gender, age, and race/ethnicity. (Race and ethnicity are defined as separate demographic characteristics, but are combined for this procedure).
- State coverage adjustment groups data within each state into cells defined by gender, age, and race.
- Second-stage ratio adjustment has three steps (national race, national ethnicity, and state) with cells that are defined using geography, gender, age, race, and ethnicity. The steps are iterated 10 times to ensure that the number of weighted-up respondents can almost exactly match a large set of external population controls. The symbol $y(t)^{SS}$ is used to indicate a simple weighted estimate for month t that uses second-stage weights, that is the weights that result after all of the weighting procedures through second-stage ratio estimation have been applied.
- A final weighting procedure, called composite weighting, is also a raking procedure with three steps (national race, national ethnicity, and state). The steps are iterated 10 times to ensure that the number of weighted-up respondents can almost exactly match a large set of external and internal controls defined by geography, gender, age, race, ethnicity, and labor force status. A subset of second-stage population controls is used in the procedure. Composite estimates of labor force status (employed, unemployed, and not-in-labor-force) that can be made from the CPS are used to split each population control into 3 controls.

## 2.3 CPS Composite Estimation

Composite estimation is a procedure that exploits month-to-month sample overlap in the CPS to lower variances on month-to-month change. The CPS sample is divided into 8 panels. In consecutive months t-1 and t, 6 panels are in common, and can be used to compute a month-to-month change estimate $\Delta(t\text{-}1,t)$. The change estimate can be used in the formula $y(t\text{-}1)^{SS} + \Delta(t\text{-}1,t)$ to update an estimate from the prior month to the current month. An example of a simple type of composite estimate is $y(t)^{C} = .6y(t)^{SS} + .4[y(t\text{-}1)^{SS} + \Delta(t\text{-}1,t)]$ which is a simple weighted average of two different estimates of the same quantity of interest. In a continuing system a more common form that would be used is $y(t)^{C} = .6y(t)^{SS} + .4[y(t\text{-}1)^{C} + \Delta(t\text{-}1,t)]$ where it is the composite estimator $y(t\text{-}1)^{C}$ from the previous month that is updated. Weighting coefficients can be optimized (.6 and .4 are just provided as an example). The composite estimator used for CPS is more complicated and has an extra term.

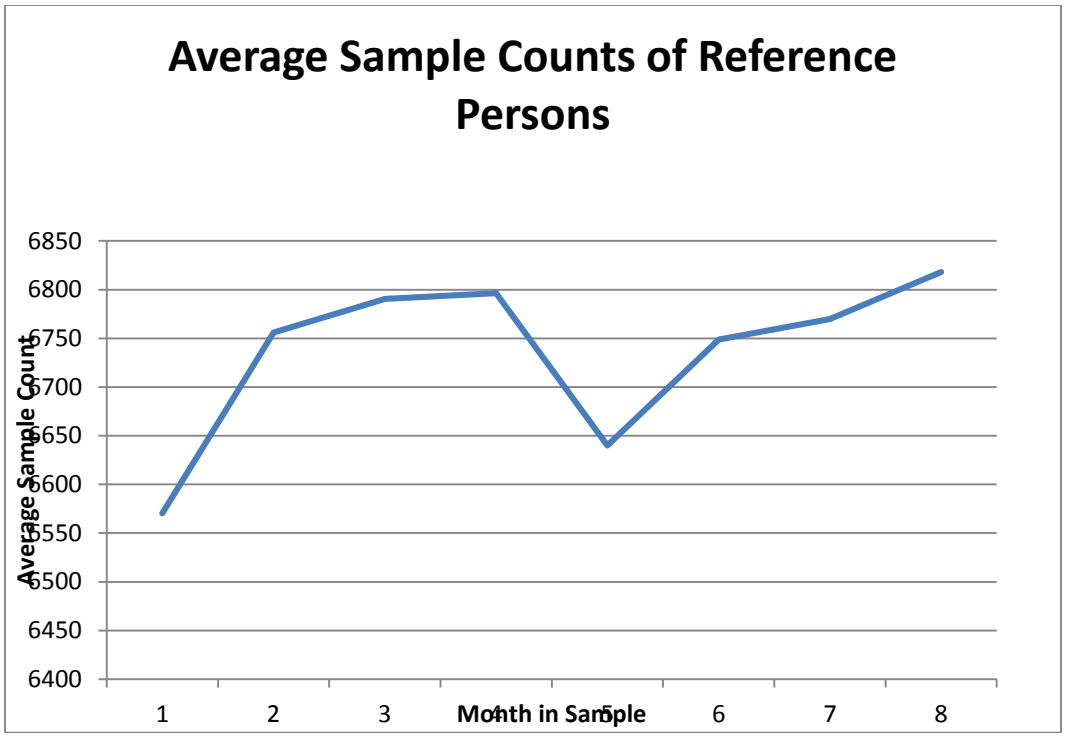
Whether a simple or a complicated form of composite estimator is used, knotty potential bias problems arise (refer to Erkens, 2012). For example, CPS has known month-in-sample bias where panel estimates consistently are different depending on how many times a particular panel has been included in the sample. The characteristics of MIS bias have changed over the years. The MIS bias affects all terms of the CPS composite estimator. The main problem is that the current compositing procedure produces labor force estimates that are systematically different from "unbiased" second-stage estimates. (It is recognized that second-stage estimates are not unbiased, but no credible statistically defensible explanation has been given for the systematic differences.) The panel attrition results presented here may well impact on MIS bias, but the research has not yet examined that possibility.

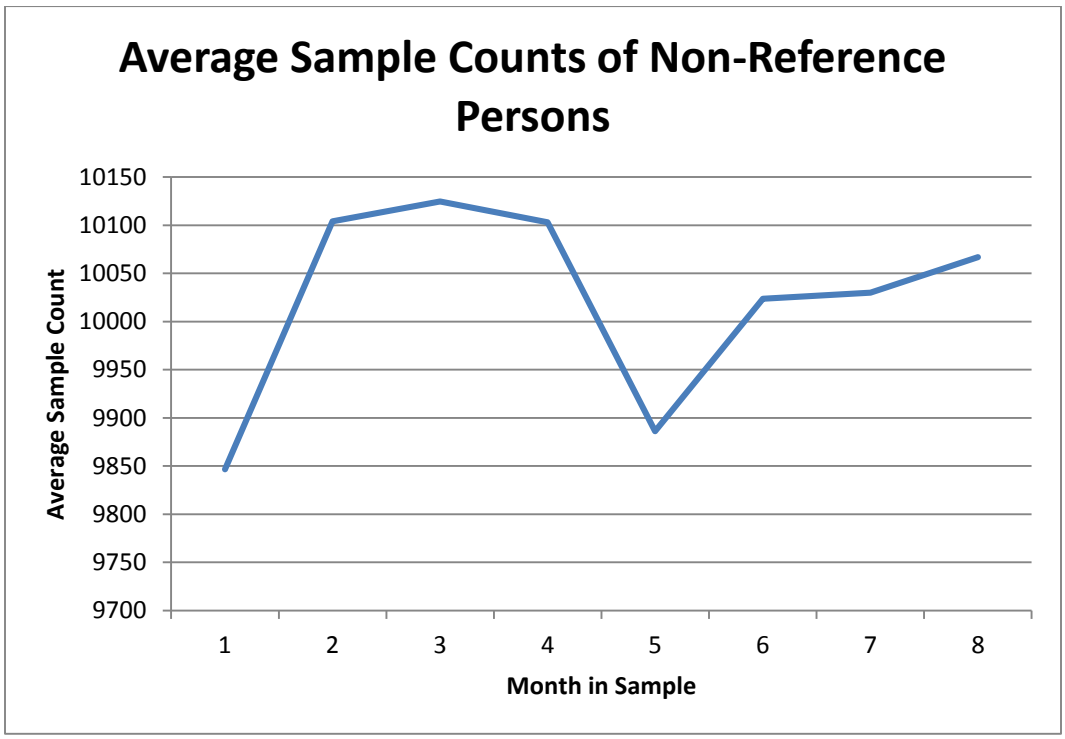
## 3. Some Research Results on Differential Response and Attrition

Several household and person variables were considered for examining differential response that might affect labor force estimates or housing unit estimates made from the CPS. After a tentative look, a person variable that looked promising was Person Type.

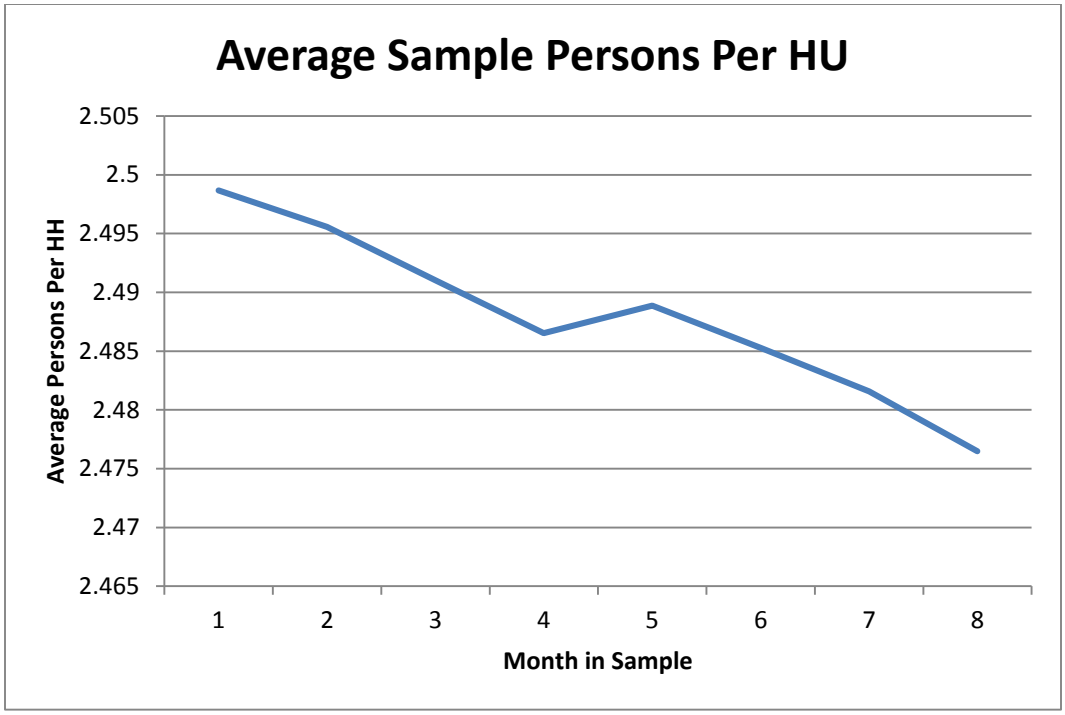
- reference person with relatives
- reference person without relatives
- spouse
- child
- grandchild
- brother/sister
- other relative of reference person
- foster child
- nonrelative of reference person, with own relatives
- nonrelative of reference person without own relatives
- unmarried partner with relatives
- unmarried partner without relatives
- housemate/roommate with relatives
- housemate/roommate without relatives
- roomer/boarder with relatives
- roomer/boarder without relatives

The average monthly sample counts (2006-2012) of reference persons and non-reference persons show the influence of different response rates by MIS.


Average Sample Counts of Reference Persons
6850
6800
6750
6700
6650
6600
6550
6500
6450
6400
Average Sample Count
1
2
3
4
5
6
7
8
Month in Sample

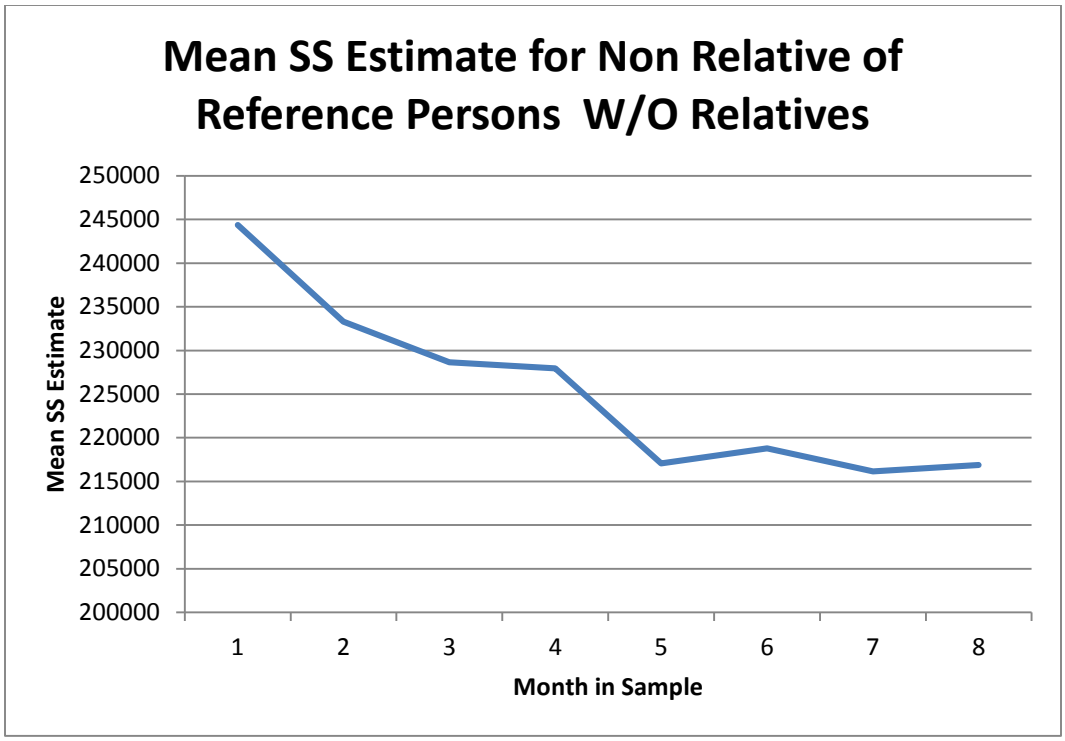

Average Sample Counts of Non-Reference Persons
10150
10100
10050
10000
9950
9900
9850
9800
9750
9700
Average Sample Count
1
2
3
4
5
6
7
8
Month in Sample

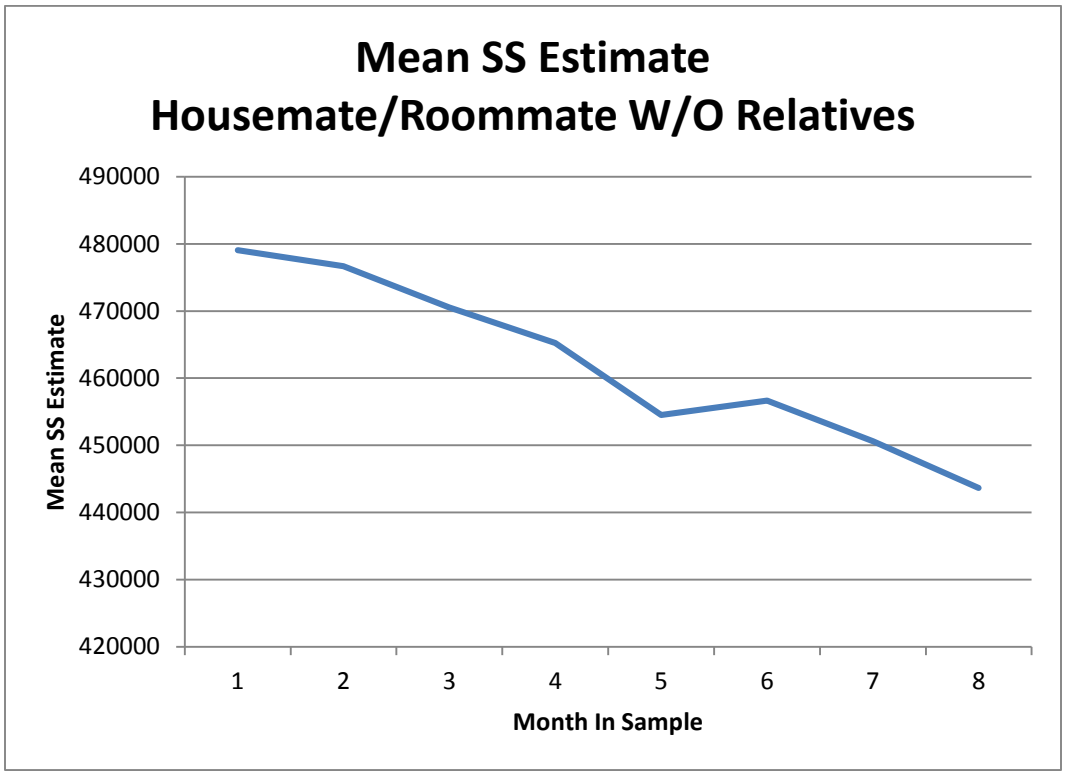
However, the simple expedient of computing an average number of reference persons per housing unit removes the influence of the different response rates. The balance clearly changes and the average declines almost 1% from about 2.499 in MIS1 to 2.476 in MIS8.




This has the appearance of mild attrition. But it is not classic attrition where a household drops out of the survey and stays out. For lack of a better term, it is a more complex attrition in probability. It can be partly due to larger households tending to fall out of the survey. It can be partly due to persons dropping off of HU rosters (without similar persons being added onto the rosters of other HUs).

The apparent attrition is much more pronounced for persons only "loosely" related to the reference person or the housing unit. The following plots show average estimates of the number of persons using second-stage weights for two person types. Estimates for "nonrelative of reference person without own relatives" fall over 11% from MIS1 to MIS8. Estimates for "housemate/roommate without relatives" fall about 8% from MIS1 to MIS8.


Mean SS Estimate for Non Relative of Reference Persons W/O Relatives
250000
245000
240000
235000
230000
225000
220000
215000
210000
205000
200000
Mean SS Estimate
1
2
3
4
5
6
7
8
Month in Sample


Mean SS Estimate Housemate/Roommate W/O Relatives
490000
480000
470000
460000
450000
440000
430000
420000
Mean SS Estimate
1
2
3
4
5
6
7
8
Month In Sample

More work is needed to quantify the possible effect on labor force characteristics of the indicated attrition in probability. However, the overall effect on HU estimates made from CPS was estimated. (Note that to estimate the number of HUs the weight of one person per sampled HU is used, usually the weight of the reference person or the reference person's spouse). The December 2012 estimate of occupied HUs using final composite weights was about 120,000,000. Assuming MIS1 is the most accurate, that estimate is about 500,000 too high. Since the same population controls are used for each MIS, the weights of reference persons tend to increase from MIS1 to MIS8 to make up for the attrition (in probability) of non-reference persons – and that increases the estimate of the number of HUs.

## 4. The Context of Housing Unit (or Household) Estimation

Part of the reason for initiating a new look at possible differential response and coverage problems in the Current Population Survey was this: a concern with discrepancies in housing unit (or household) estimates that are made from several surveys conducted by the Census Bureau and the 2010 demographic Census (Cresce, 2003). The different estimates of HUs are not easily explained by factors such as coverage differences, data collection differences including different mixes of modes, processing differences, etc. For example, here are a few estimates of occupied HUs from 2010. The CPS estimate of households is largest, even though it should logically be lower than the ACS due to the coverage restriction to the Civilian Noninstitutional Population.

- 117,538,000 occupied HUs Current Population Survey (CNP)
- 116,716,000 occupied HUs 2010 demographic census (resident population)
- 114,567,000 occupied HUs American Community Survey (resident population)

Here are highly rounded December 2012 estimates of occupied HUs from different stages of CPS weighting:

- 96 Million using base weights of respondents r
- 106 million housing unit weighting cell nonresponse adjustment (r+Type_A)/r
- 106 million first-stage ratio adjustment
- 120 million national and state coverage adjustment
- 120 million second-stage ratio adjustment
- 120 million composite weighting

### 4.1 HU Estimates After Nonresponse Adjustment

The initial weighting steps of base weighting and nonresponse adjustment are housing-unit based. Ideally, housing-units weighting procedures would result in a good estimate of the number of occupied in-scope HUs. Frame coverage problems are possible, but per ongoing Census research frame noncoverage problems are minimal (although CPS and some other surveys are participating in coverage improvement projects). Survey procedures drive the jump from the basic-weighting 96 million HUs to the 106 million HUs after nonresponse adjustment; the procedures are another possible source of error.

A jump in occupied HU estimates due to nonresponse adjustment is expected. The jump from 96 million HUs to 106 million depends on the proper identification of Type_A HUs that are occupied and in scope, but that do not respond. The frame for the survey also includes:

- Type_B HUs that are temporarily out-of-scope but could be in-scope in following months, such as vacant houses

- Type_C HUs that are permanently out-of-scope, such as destroyed houses.

The frame has no indication of Type_B or Type_C and that determination is made only on sample cases using CPS procedures. Some approximate 2012 counts follow.

- N = 135,000,000 frame HUs including occupied in-scope and out-of scope
- n = 77,000 "designated" sampled HUs monthly
- r = 53,500 responding HUs
- Type_A 6,500 occupied in-scope HUs not responding
- Type_B 12,500 temporarily out-of-scope (so no response should be obtained)
- Type_C 4,500 permanently out-of-scope (so no response should be obtained)

There is no evidence that CPS procedures are in error, so the comment here is just speculative. About ¾ of Type_B HUs are simply categorized as vacant, but if half of those were really Type_A the estimate of HUs after nonresponse adjustment would jump up an additional 8,000,000 to 114,000,000 occupied in-scope HUs. Although not a differential response issue, CPS panel data could be used to identify probable false Type_B determinations. For example, if the same housing unit with the same people responds in March and the following May but not in April, a Type_B determination in April would not be expected.

## 4.2 HU Estimates After Person Weighting (Population Controlling)

CPS coverage of persons varies by gender, age, race, and ethnicity. The two coverage steps, second-stage ratio adjustment, and composite weighting all use population controls. CPS weights are adjusted so that weighted sample data is benchmarked to those controls. Respondents may not resemble nonrespondents, so any bias due to differential coverage by gender/age/race/ethnicity may be reduced but is not eliminated. The population controls are developed by Census and are external to CPS (US Census Bureau, 2006). There is no evidence of major upward/downward bias that would also bias CPS estimates of housing units.

The CPS coverage of adults in December 2012 was about .88 – the ratio of adults estimated after nonresponse adjustment to the population control total of all adults. Note that this is very close to the implied coverage of households given by the ratio 106,000,000/120,000,000. Person weights are increased to match population controls, and that causes the increase of estimated CPS occupied households from 106,000,000 after nonresponse adjustment to 120,000,000 after population controlling.

Note that even if the population controls were discovered to be high/low and corrected, the two ratios would remain about the same using current CPS methodology. It does not follow that the household ratio must be that close to the person coverage ratio.

- CPS already has corrections for one type of differential household response in nonresponse adjustment, but it is based strictly on geography. Other corrections for differential household response based on type of household could affect the ratio.
- CPS already has corrections for some differential person coverage through population controls. Other corrections for person coverage could affect the ratio. Section 3 indicates the 120,000,000 HUs after population controlling may be 500,000 to high, based on differential coverage (or attrition-in-probability) by MIS of non-reference persons.

### 4.3 The Importance of Getting the Number of Persons Listed Right

One aspect of this is the importance of getting the proper number of persons listed. The person coverage ratio of .88, naively interpreted, indicates that too few persons are listed. There are three possible causes of "listing too few persons" that can be readily pointed out.

- Listings are okay, but the HU mix is wrong --Perhaps HUs with larger numbers of persons respond with less frequency than smaller housing units. This is a hypothetical suggestion only and has not been investigated, but the panel nature of CPS would allow an examination of differential coverage of household size by MIS.
- Listings are initially okay, but there is attrition -- Perhaps some person types are systematically dropped from HUs from MIS1-MIS8, without similar persons being added to HUs. The finding regarding MIS attrition in probability of relationship to reference person could be due to a combination of attrition and the HU mix.
- Listings systematically have too few non-reference persons. (Any responding household has a reference person.) Internal CPS data cannot be used to look into this.

If too few persons are listed, the resulting HU estimates will be too large. Person estimates, due to population controlling, will end up at the right level. However, the necessary upward weight adjustments will also raise HU estimates. In the extreme case, the estimated 106,000,000 occupied HUs after nonresponse adjustment is nearly unbiased, and the large increase to 120,000,000 occupied HUs after population controlling is wrong.

## 5. Conclusions and Suggestions for Future Research

BLS and Census have a long-standing interest in identifying and, if possible, correcting for differential response and differential coverage. An uncorrected differential is assumed to result in bias. Evidence was given in section 3 of early research results, using relationship to reference person, indicating "attrition in probability" across the 8 months-in-sample (MIS) when HUs are in the Current Population Survey. Indications are that the December 2012 estimates of 120,000,000 occupied HUs is about 500,000 too high as a result. It is intended to estimate the effect on topside labor force estimates.

The research done and the extra proposed here is limited to topics that can be examined using CPS data without reference to other surveys. The panel structure of the CPS can be exploited, housing units being in sample 8 months. The research findings given here were based only on analysis by month-in-sample, but microdata could be linked. Possibilities include:

- Measuring differential response by housing unit size and other household variables
- Determining if the attrition linked to relationship to reference person is due to listing problems or other causes
- Link HU data across the 8 MIS and look for unlikely Type_B vacant determinations. (ex: The HU responds in March, in April it is determined to be a Type_B vacant and temporarily out-of-scope, and the HU responds in May.)

The Current Population Survey is a labor force survey, and estimating housing units is not a top priority. In this paper information was given on the estimation of housing units at the different stages of CPS weighting. For December 2012 occupied housing unit estimates there is an increase from 106,000,000 after housing unit nonresponse to 120,000,000 after person weighting using population controls. Although on the high side, CPS HU estimates after person weighting are closer to estimates from other Census surveys than CPS HU estimates made after HU nonresponse adjustment.

- It is my opinion that a good goal is to get the HU estimates right during HU weighting, and a secondary related goal is for person weighting to have minimal effect on HU estimates.
- I think it is likely that improvements to HU estimation would also lead to modest improvement in labor force estimation.

## References

Cresce, Arthur R., Yang Cheng, and Christopher Grieves. 2013. Household Estimates Conundrum: Effort to Develop More Consistent Household Estimates Across Surveys. Paper presented at the 2013 meeting of the Federal Committee on Statistical Methodology. Washington, DC.

Erkens, Greg. 2012. Changes in Panel Bias in the U.S, Current Population Survey and its Effects on Labor Force Estimates. *JSM Proceedings*, Section on Survey Research Methods. Alexandria, VA; American Statistical Association.

Robison, Edwin and Martha Duff. 2003. Redesigned Current Population Survey Raking. *JSM Proceedings*, Section on Survey Research Methods. Alexandria, VA; American Statistical Association.

US Census Bureau. 2006. *Current Population Survey Design and Methodology, Technical Paper 66*. (www.census.gov/cps/methodology/techdocs.html)