

# Impact of MAF-Based Frame Coverage on Survey Estimates

Xijian Liu<sup>1</sup>

US Census Bureau

4600 Silver Hill Road, Washington, DC 20323

## Abstract

One of the goals of the 2010 Demographic Survey Sample Redesign is to consider using the continually updated Master Address File as the primary source to develop the sampling frames for current demographic household surveys. To support this goal, the Census Bureau conducted several coverage evaluations and found that coverage differences existed between the current frames and the MAF-based frame. This paper investigates how Current Population Survey estimates are impacted if current permit and area frame sample addresses not on the MAF-based frame are excluded.

**Key Words:** Coverage Errors, Master Address File, Current Population Survey

## 1. Background

The U.S. Census Bureau currently uses a four-frame system to support several major demographic household surveys. These surveys include the Current Population Survey (CPS), the Survey of Income and Program Participation (SIPP), the Consumer Expenditure Surveys (CE), the American Housing Survey (AHS), the National Crime and Victimization Survey (NCVS), and the State Children Health Insurance Program (SCHIP). The existence of the Master Address File (MAF) provides a simpler and more cost efficient frame option for these surveys. The U.S. Census Bureau is currently conducting evaluations to determine what would be the impact on household surveys if we use the MAF as the sole source of addresses to construct the sampling frames for the 2010 Demographic Surveys Sample Redesign (DSSR).

The universe of interest for most of the current household surveys consists of the housing unit population and the non-institutional non-military group quarter population in the U.S. for every survey reference date. For some surveys, the universe of interest consists housing units instead of persons. These surveys publish a wide range of social and economic data, including data for employment and labor force, income and welfare, consumer expenditure, housing, and crime statistics. Most surveys publish data for the U.S. and for various demographic groups. The CPS and SIPP also publish data at the state level.

A common frame system is used to cover housing units and group quarters in the U.S. to meet the needs of multiple surveys. Currently, the common frame system consists a unit frame, a group quarters (GQ) frame, a permit frame, and an area frame. In most areas of the U.S., the unit frame and the GQ frame capture housing units and group quarters which existed at the time of Census 2000; while the permit frame captures new housing units added after Census 2000. The sources of addresses for these three frames are the address lists from Census 2000 and from building permits. In other areas that are mostly rural in nature, an area frame is used. This final frame uses addresses from ongoing field

<sup>1</sup> Xijian Liu is a mathematical statistician in the Demographic Statistical Methods Division of the U.S. Census Bureau. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical and methodological issues are those of the author and not necessarily those of the U.S. Census Bureau.

listings. The current area frame covers approximately 12% of the addresses in the U.S. The field operations for developing the area frame and the permit frame are very expensive.

The MAF is the U.S. Census Bureau's national inventory of addresses for all known living quarters. Operations associated with decennial censuses provide major updates to the MAF every ten years. The U.S. Census Bureau also updates the MAF using the Delivery Sequence File (DSF) from the U.S. Postal Service every six months. There are various operations that update the MAF at a smaller scale as well. The American Community Survey (ACS) uses the MAF to construct its sample frame.

From 2004 to 2008, the U.S. Census Bureau conducted several evaluations to gain understanding of the coverage differences of the two frame systems. A series of reports and papers from these evaluations are given by Flanagan et al. (2007), Kennel (2007), Kennel and Corlett (2005), and Loudermilk and Li (2009). These studies showed that coverage differences of the two frame systems are relatively small at the national level (Liu, 2008).

The main differences of the current frame system and the MAF-based frame are in the sub-universes currently covered by the permit frame and the area frame. In the permit frame sub-universe, the current frame system uses the building permits and the MAF-based frame uses the DSF. The current area frame uses a dependent listing operation starting from a list of addresses from the recent MAF. Listers can add units found on the ground that were not on the MAF. The MAF-based frame on this paper is a frame developed using the MAF as the sole source and without area listing. Based on our evaluations, units on the current permit frame and area frames but not found on the MAF-based frame were approximately 1.3% and 1.7% of all addresses in the entire U.S. in 2007 respectively. The question is what will be the impact of this coverage difference on survey estimates.

This paper will investigate the impact on selected labor force estimates if addresses not found on the MAF were excluded from the existing CPS samples. The study using CPS data can also give us some information about the potential impact on estimates of other surveys. In a broader context, this study can provide some insights into how new construction coverage and rural coverage can impact survey estimates as well.

This paper is a short version of a U.S. Census Bureau internal report by Liu (2009).

## 2. Coverage Bias

The estimates we consider are Civilian Labor Force Participation (CLF) rates and Unemployment (UE) rates for the U.S. and for the fifty states and District of Columbia. We will investigate the changes to the estimates when permit frame and area frame units not on the MAF-based frame are excluded from the existing samples. If we assume the CLF rates and UE rates derived from the existing samples are unbiased, then the changes to the estimates are the coverage bias due to the MAF-based frame undercoverage.

Frame coverage bias for means can be expressed as the product of the percent of population elements not covered by the frame and the difference of the mean over the covered elements and the mean over the non-covered elements (Lessler and Kalsbeek,

1992). Let  $\bar{Y}$  be the mean over all population elements. Let  $\bar{Y}_A$  and  $\bar{Y}_0$  be the mean over population elements covered by the frame and the mean over population elements not covered by the frame respectively. Let  $W_0$  be the proportion of population elements not covered by the frame. Then

$$\bar{Y}_A - \bar{Y} = W_0(\bar{Y}_A - \bar{Y}_0).$$

If the coverage difference between the current four-frames and the future MAF-based frame is small, the impact on survey estimates would be small unless the MAF-based frame omits units that have very different characteristics.

The data we use for this study are the November 2005, March 2006, and July 2006 CPS samples. An address, if selected, is in CPS sample for four consecutive months, out-of-the sample for eight months, and in sample for another four months. Because these three CPS samples we chose are four months apart, they do not contain common sample addresses. These three monthly samples contain more than 300,000 responded person units. Note that although the CPS selects address samples from an address frame, the ultimate units of interest for this survey are persons.

In this paper, CPS sample addresses that were omitted from the MAF-based frame are referred to as non-matches. The area frame non-matches are CPS area frame sample addresses that would not be on the MAF-based frame without an area listing operation, i.e., the addresses added by the listing operations. Permit frame non-matches are CPS permit frame sample addresses that were not on the MAF or entered the MAF too late for survey sampling and interviewing, i.e., first appearance on the MAF was less than four months before interviewing. Table 2.1 shows the percent of CPS persons living at non-matched addresses. These are smaller than the 1.3% and 1.7% for 2007 mentioned earlier. The lower percentage is mainly resulted from the time difference and the exclusion of addresses not yet built and vacant addresses.

**Table 2.1:** Persons at Non-Matched Addresses

	Number of Person Units	Persons at Non-Matched Addresses as Percent of Persons in the Entire Sample	
		Permit Frame	Area Frame
Nov 2005	106,318	1.14%	0.90%
Mar 2006	104,194	1.09%	0.86%
Jul 2006	105,960	0.99%	0.97%

The persons at area frame non-matched addresses account for a larger proportion of the sample units in some states. For example, persons at area frame non-matched addresses were 4.2%, 5.1%, and 4.8% of the persons in CPS sample in West Virginia for the three months under consideration respectively. Throughout the paper, standard errors are shown in parentheses in tables.

Table 2.2 and Table 2.3 show clear differences of CLF rate and UE rate for permit frame non-matched addresses and for addresses in the entire four-frames. Persons living at newly built addresses appear more likely to be in labor force and less likely to be unemployed than persons living at a typical addresses in the four-frames. However, there is no evidence that persons living at permit frame non-matched addresses have different CLF rate or UE rate than persons living at permit frame matched addresses.

**Table 2.2:** CLF Rates in Percent for Matched and Non-Matched Addresses

	3-month Average	Nov 05	March 06	Jul 06
Four-Frames	66.5 (0.1)	66.3 (0.2)	65.9 (0.2)	67.3 (0.2)
Permit Frame				
Permit Frame Matches	72.6 (0.4)	72.4 (0.9)	71.5 (0.7)	73.8 (0.7)
Permit Frame Non-Matches	72.4 (1.6)	73.4 (1.8)	73.1 (2.4)	70.4 (2.5)
Area Frame				
Area Frame Matches	61.8 (0.4)	61.8 (0.7)	61.0 (0.6)	62.7 (0.7)
Area Frame Non-Matches	64.0 (1.3)	62.1 (2.1)	61.8 (2.5)	67.9 (2.0)

**Table 2.3:** UE Rates in Percent for Matched and Non-Matched Addresses

	3-month Average	Nov 05	March 06	Jul 06
Four-Frames	4.9 (0.1)	4.9 (0.1)	4.9 (0.1)	5.0 (0.1)
Permit Frame				
Matches	3.3 (0.2)	3.1 (0.3)	3.2 (0.3)	3.5 (0.3)
Non-Matches	3.1 (0.4)	3.4 (0.7)	2.8 (0.8)	3.0 (0.7)
Area Frame				
Matches	4.8 (0.2)	4.4 (0.3)	5.0 (0.3)	5.1 (0.3)
Non-Matches	4.9 (0.6)	5.7 (1.3)	4.1 (0.9)	4.9 (0.9)

Table 2.2 and Table 2.3 also show some evidence of differences of CLF rate between area frame non-matched addresses and a typical address in the entire four-frames. Persons living in rural areas appear less likely to be in labor force than persons living in a typical address in the four-frames.

We now take a closer look at the UE rates for area frame addresses. Persons living in mobile homes consisted almost one fifth of the CPS sample person units in the area frame and approximately one third of the sample person units at area frame non-matched addresses. As shown in Table 2.4, UE rates were higher for persons living at area frame non-matched addresses of mobile homes than for persons living in a typical address in the four-frames. Omitting these mobile homes in the area frame could cause downward bias to UE rate estimates. On the other hand, UE rates were lower for persons living in area frame non-matched addresses of conventional housing units than for persons living in a typical address in the four-frames. Omitting these conventional housing units in the area frame could cause upward bias to UE rate estimates. Using state estimates as an example, the combined effect of omitting mobile homes and omitting conventional housing units in the area frame can cause bias to state UE rates in either direction depending on the proportion of mobile homes among the omitted addresses in a particular state.

**Table 2.4:** UE Rates in Percent for Mobile Homes and Conventional Housing Units in the Area Frame

	3-month Average	Nov 05	March 06	Jul 06
Four-Frames	4.9 (0.1)	4.9 (0.1)	4.9 (0.1)	5.0 (0.1)
Area Frame Mobile Homes				
Matches	7.4 (0.9)	6.4 (0.9)	8.6 (0.9)	7.2 (1.0)
Non-matches	8.6 (1.2)	9.8 (2.4)	9.0 (2.3)	7.1 (1.7)
Area Frame Conventional HUs				
Matches	4.3 (0.2)	4.0 (0.3)	4.3 (0.3)	4.7 (0.3)
Non-matches	2.7 (0.6)	2.6 (1.1)	1.8 (0.7)	3.8 (1.1)

So far, the discussion in this section has shown that omitting the non-matched addresses in the permit frame and in the area frame would cause non-zero bias some CLF or UE rate estimates. The question is whether the magnitude of bias is big enough for concerns.

We exclude the non-matches from the CPS samples, rerun the CPS weighting process, and calculate new CLF rates and UE rates using the new weights. Then we compare the new estimates to the original estimates derived from the full sample. Standard errors are computed using the associated replicate weights. For more details about the CPS weighting process, the readers can refer to CPS Technical Paper 66 (2006).

The following definitions and symbols will be used throughout this paper. Let  $D$  be the difference of the reweighted new estimate and the original estimate (for a single month). If the original estimates are unbiased, then the difference can be used to measure the coverage bias. The net bias  $B$  is estimated using the three-month average value of  $D$ . Let  $X$  and  $V$  be the CLF rate or UE rate estimate and variance. Then the relative bias  $RB$  is estimated using the three-month average of  $D/X$ . The ratio of squared net bias to MSE,  $R$ , is estimated using the three-month average of  $D^2/(V+D^2)$ . For more about net bias, relative bias, and the ratio of squared bias to MSE, the readers can refer to Lessler and Kalsbeek (1992).

### 3. Impact on US Estimates

Addresses on the current permit frame and area frame but not on the MAF-based frame were a small percent of the entire U.S. address frame. In the previous section, we showed that omitting these addresses from the address frame would cause non-zero bias to some CLF or UE rate estimates. In this section, we investigate whether the magnitude of bias to U.S. estimates is large enough for concerns.

Table 3.1 and 3.2 show the changes to the CLF rate and UE rates for the U.S. An asterisk indicates the 90% confidence interval for the change estimate does not contain zero.

**Table 3.1:** Change to CLF Rate in Percent for the U.S.

	Original Estimate	Change When Non-Matches Were Removed ( $D$ )	
		From the Permit Frame	From the Area Frame
Nov 2005	66.33 (0.15)	0.00 (0.04)	+0.05 (0.04)
March 2006	65.92 (0.16)	-0.02 (0.01)	+0.04* (0.02)
July 2006	67.25 (0.15)	+0.01 (0.02)	+0.01 (0.01)

**Table 3.2:** Change to the UE Rate in Percent for the U.S.

	Original Estimate	Change When Non-Matches Were Removed ( $D$ )	
		From the Permit Frame	From the Area Frame
Nov 2005	4.93 (0.11)	+0.02 (0.03)	-0.01 (0.03)
March 2006	4.86 (0.09)	+0.02 (0.01)	0.00 (0.01)
July 2006	5.02 (0.09)	+0.02* (0.01)	0.00 (0.01)

Notice that the changes in the area frame column of Table 3.1 and the permit frame column of Table 3.2 are all positive.

Table 3.3 shows the estimated net bias, relative bias, and the ratio of squared net bias to MSE for U.S. CLF rate and UE rate estimates.

**Table 3.3: Bias Estimates in Percent for the U.S. CLF Rate and UE Rate**

	Estimated Bias for CLF Rates When Non-Matches Were Removed		Estimated Bias for UERates When Non-Matches Were Removed	
	From the Permit Frame	From the Area Frame	From the Permit Frame	From the Area Frame
<i>B</i>	0.00 (0.015)	0.03* (0.015)	0.02* (0.010)	0.00 (0.010)
<i>RB</i>	0.00	0.05	0.35	-0.03
<i>R</i>	0.1	11.8	8.9	0.1

These results confirm that there would be a non-zero bias for the U.S. CLF rate due to area frame non-matches and for the U.S. UE rate due to permit frame non-matches. However, the magnitudes of these biases are small.

Similar analyses were performed for domain estimates for race sex age groups and ethnicity sex age groups. The results also showed at most a small amount of bias to these estimates (Liu, 2009). In the next section, we will focus on estimates at the state level.

#### 4. Impact on State Estimates

In this section, we investigate whether the magnitude of bias to state estimates is large enough for concerns.

As a result of the sample design changes in January 1996, CPS no longer produces “direct state estimates” for the 11 largest states. The state monthly sample size for CPS is not designed to produce monthly state estimates with sufficient reliability. CPS uses a model based on a signal-plus-noise approach (CPS Tech paper 66, 2006). For this study, the state estimates of CLF rate and UE rate are the three-month average rates using the November 2005, March 2006, and July 2006 state samples.

We first look at the effect of excluding the permit frame non-matches. The average permit frame non-match rates over the three interview months vary among the 50 states and District of Columbia. Table 4.1 shows the states and the corresponding ranges of non-match rate.

**Table 4.1: Average Permit Frame Non-Match Rate for States**

Non-Match Rate	State
0.00-0.99%	AL, CA, CT, DC, HI, IA, IL, KS, KY, MA, MI, MO, MT, ND, NE, NJ, NY, OH, OK, PA, RI, TN, WV
1.00-1.99%	AK, AR, AZ, DE, FL, GA, IN, LA, MD, ME, MN, MS, NC, NH, NV, OR, SC, TX, VA, VT, WA, WI, WY
2.00-2.99%	CO, ID, NM, SD, UT

Table 4.2 and Table 4.3 show the average changes of state estimates for CLF rate and UE rate when permit frame non-matches were excluded. The average is taken over the states within each non-match rate category.

Notice that the columns for the average of  $B$  are all negative in Table 4.2 and all positive in Table 4.3. Also, there is a clear trend that the size of the differences increases as the non-match rate increases.

**Table 4.2:** Average Change Estimates for State CLF Rate Estimates  
Due to Permit Frame Non-Matches

Non-Match Rate	Number of States	Average $B$ In Percent	Average $RB$ In Percent	Average $R$ In Percent
0.00-0.99%	23	-0.02 (0.03)	-0.03	0.67
1.00-1.99%	23	-0.03 (0.03)	-0.04	1.43
2.00-2.99%	5	-0.08 (0.07)	-0.12	2.42
All States and DC	51	-0.03 (0.02)	-0.04	1.18

**Table 4.3:** Average Change Estimates for State UE Rate Estimates  
Due to Permit Frame Non-Matches

Non-Match Rate	Number of States	Average $B$ In Percent	Average $RB$ In Percent	Average $R$ In Percent
0.00-0.99%	23	0.01 (0.02)	0.25	0.49
1.00-1.99%	23	0.02 (0.02)	0.48	1.54
2.00-2.99%	5	0.07* (0.03)	1.74	2.18
All States and DC	51	0.02* (0.01)	0.50	1.13

Table 4.4 and Table 4.5 show the largest absolute values for change estimates in each non-match rate category. The bias measures in these two tables are relatively small.

**Table 4.4:** Maximum Change Estimates in Percent  
for State CLF Rate Due to Permit Frame Non-Matches

Non-Match Rate	$max B $	$max RB $	$maxR$
0.00-0.99%	0.14	0.19	1.93
1.00-1.99%	0.22	0.30	4.01
2.00-2.99%	0.32	0.44	7.12

**Table 4.5:** Maximum Change Estimates in Percent  
for State UE Rate Due to Permit Frame Non-Matches

Non-Match Rate	$max B $	$max RB $	$maxR$
0.00-0.99%	0.04	1.20	1.53
1.00-1.99%	0.11	3.05	9.14
2.00-2.99%	0.11	2.63	4.04

Next, we repeat the above state level analysis for the area frame.

**Table 4.6:** Average Area Frame Non-Match Rate for States

Non-Match Rate	State
0.00-0.99%	AZ, CA, CO, CT, DC, DE, FL, HI, IA, ID, IL, IN, KS, LA, MA, MD, MI, ND, NE, NH, NJ, NV, NY, OH, OR, PA, RI, TN, UT, VA, WA, WI
1.00-1.99%	GA, MN, MO, NC, SC, SD
2.00-2.99%	AL, ME, TX, VT, WY
3.00-3.99%	AK, KY, OK
4.00-4.99%	AR, MS, MT, NM, WV

Table 4.7 and Table 4.8 show the summary of state level change estimates for CLF rate and UE rate when area frame non-matches were excluded.

**Table 4.7:** Average Change Estimates for State CLF Rate Due to Area Frame Non-Matches

Non-Match Rate	Number of States	Average <i>B</i> In Percent	Average <i>RB</i> In Percent	Average <i>R</i> In Percent
0.00-0.99%	32	+0.02 (0.01)	0.03	0.21
1.00-1.99%	6	+0.08 (0.07)	0.12	1.84
2.00-2.99%	5	+0.07 (0.06)	0.11	3.00
3.00-3.99%	3	+0.25 (0.18)	0.38	6.40
4.00-4.99%	5	+0.08 (0.15)	0.11	5.22
All States and DC	51	+0.05* (0.02)	0.08	1.53

All measures on the column of average *B* in Table 4.7 are positive. Unlike Table 4.3, Table 4.7 does not show a very clear pattern of increasing magnitude of change for state CLF rate as the non-match rate increases.

**Table 4.8:** Average Change Estimates for State UE Rate Due to Area Frame Non-Matches

Non-Match Rate	Number of states	Average <i>B</i> In Percent	Average <i>RB</i> In Percent	Average <i>R</i> In Percent
0.00-0.99%	32	+0.00 (0.01)	0.06	0.17
1.00-1.99%	6	-0.02 (0.03)	-0.42	1.96
2.00-2.99%	5	+0.01 (0.03)	-0.08	2.21
3.00-3.99%	3	+0.03 (0.14)	0.61	1.33
4.00-4.99%	5	-0.11 (0.09)	-2.22	7.46
All States and DC	51	-0.01 (0.01)	-0.20	1.36



On the column of average  $B$  in Table 4.8, there are some positive and negative measures. However, missing the area frame non-matched addresses can change the UE rates in either direction.

**Table 4.9:** Maximum Change Estimates in Percent for State CLF Rate Due to Area Frame Non-Matches

Non-Match Rate	$\max  B $	$\max  RB $	$\max R$
0.00-0.99%	0.10	0.14	1.88
1.00-1.99%	0.27	0.40	7.12
2.00-2.99%	0.18	0.26	5.43
3.00-3.99%	0.46	0.73	7.58
4.00-4.99%	0.42	0.61	11.56

**Table 4.10:** Maximum Change Estimates in Percent for State UE Rate Due to Area Frame Non-Matches

Non-Match Rate	$\max  B $	$\max  RB $	$\max R$
0.00-0.99%	0.03	0.76	1.11
1.00-1.99%	0.12	2.18	5.73
2.00-2.99%	0.04	1.39	3.73
3.00-3.99%	0.10	1.62	1.66
4.00-4.99%	0.49	8.78	22.54

To interpret the results of Table 4.10, the readers should keep in mind that the UE rates in this period were approximately five percent. The larger bias measures in the bottom row of Table 4.10 came from a large difference of UE rate for West Virginia. For this state, we observed a moderate amount of bias for the state UE rate estimate.

## 5. Conclusions and the Next Steps

Our analysis showed that permit frame and area frame addresses omitted from the MAF-based frame had different labor force characteristics. Missing these addresses can cause non-zero coverage bias to labor force estimates. Our analysis also showed that for most of the estimates published for CPS, the sizes of coverage bias would not be large enough to cause concerns. However, the risk of non-trivial coverage bias for state level estimates exists in some states that have a large proportion of addresses in rural areas.

To help interpret the permit frame results from this study, we should point out that the MAF-based frame contains post-2000 newly built addresses that are not on the permit frame (Flanagan, 2008). If the labor force characteristics of persons living at these addresses and the labor force characteristics of persons living at permit frame addresses not on the MAF-based frame are similar as suggested by Table 2.2 and 2.3, the actual impact on labor force estimates due to the new construction coverage differences of the current frame and the MAF-based frame may be smaller than what were shown in this paper.

The U.S. Census Bureau is currently looking into options to deal with the potential coverage bias due to MAF undercoverage, especially rural undercoverage. Considerations will be given to quality and cost effectiveness. Then, we will make a final frame decision for the 2010 Sample redesign with survey sponsors.

## 6. References

- Flanagan, T., Loudermilk, C., and Liu, X., Analytical Report for the Frame Assessment for Current Household Surveys: New Construction Study of 2008, U.S. Census Bureau Internal Memorandum, 2008.
- Kennel, T., Second Analytical Report for Frame Assessment for Current Household Surveys: Area Frame Study, U.S. Census Bureau Internal Memorandum, 2007
- Kennel, T. and Corlett, M., Comparing the US Census Bureau's Master Address File to the Survey of Income and Program Participation Area Listings. Proceedings of the Joint Statistical Meeting, 2005
- Lessler, J. and Kalsbeek, W., Nonsampling Errors in Surveys, John Wiley & Sons, Inc., 1992.
- Loudermilk, C. and Li, M., A National Evaluation of Coverage for a Sampling Frame Based on the Master Address File, Joint Statistical Meeting, 2009
- Liu, X., Comparing the Quality of the Master Address File and the Current Demographic Household Surveys' Multiple Frames, 2007 Federal Committee on Statistical Methodology research Conference, 2007
- Liu, X., A Supplemental Report of the Frame Assessment for Current Household Surveys: Effects of MAF-Based Frame Coverage on Survey Estimates, US Census Bureau Internal Memorandum, 2009
- U. S. Census Bureau, Design and Methodology, Current Population Survey Technical Paper 66, 2006