A Preliminary Evaluation of the Residence History Data in the 2010 and 2011 Field Tests of the Reengineered Survey of Income and Program Participation

Matthew Marlay and Peter Mateyka

U.S. Census Bureau
matthew.c.marlay@census.gov
peter.mateyka@census.gov

Abstract

The Survey of Income and Program Participation (SIPP) collects longitudinal data on a variety of topics including employment and program participation. The strength of SIPP is that respondents are followed for two to four year periods, even when they change residences, enabling researchers to perform analyses not possible with cross-sectional Census surveys. The longitudinal design of SIPP means collecting accurate residence history information is vital to the success of the survey.

The Bureau is redesigning the SIPP survey instrument, and this redesign will have implications for how the survey collects residence history information. The current in production SIPP has a respondent recall period of four months and respondents are only allowed to report one move during each four month interview period, or wave. Detailed information about previous residences is collected early in the survey (Wave 2), but this information is not concurrent with data collected on respondents in later waves. The reengineered SIPP instrument uses an Event History Calendar (EHC) format with a respondent recall period of one year. Residence history information for the year is collected concurrently with data on jobs, program participation, and health insurance. Respondents can report living in up to five residences during the reference year of the survey, and detailed address information is collected for each residence reported.

We analyze residence history data collected in the 2010 and 2011 field tests of the Re-Engineered Survey of Income and Program Participation (SIPP-EHC). We use two methods to evaluate the quality of the SIPP-EHC residence history data. First, we match the addresses collected in the 2011 field test of SIPP-EHC to spatial extracts from the Census Bureau's Master Address File/TIGER database. We report the percentage of addresses successfully matched at the place, county, and state level. Second, we compare the monthly mover rates reported in the test 2011 SIPP-EHC (covering calendar year 2010) to the 2010 SIPP-EHC (covering calendar year 2009) to the 2008 SIPP (covering calendar year 2009). We primarily focus on whether there are time trends in the likelihood of respondents reporting moving in the two SIPP-EHC field tests, indicating memory decay or misreporting.

Introduction

The Census Bureau is currently reengineering the Survey of Income and Program Participation (SIPP). Like the current SIPP (called production SIPP), the primary objective of the Re-Engineered SIPP (called SIPP-EHC) is to collect accurate and comprehensive information on topics such as income, program participation, employment, and health insurance, with the additional goal of collecting this information in a more cost effective and less burdensome way. These improvements are achieved by shifting the four-month recall period of current production SIPP to a one-year reference period and introducing an event history calendar (EHC) to aid respondent recall of retrospective history information.

The purpose of this paper is to compare residence history data collected in the 2010 and 2011 SIPP-EHC field tests to data from the 2008 production SIPP. The collection of residence history information underwent several significant changes in SIPP-EHC. Most significantly, in production SIPP respondents are only allowed to report one move per four-month reference period. This move is, presumably, to the current residence. In SIPP-EHC residence history data is collected in the EHC, where the starting month and ending month (spells) of events are recorded on a visual calendar to aid respondent recall. Respondents are asked to report up to five residences they lived in during

the reference year. For each residence, respondents are asked to provide detailed address information. This represents a significant change from production SIPP, because identifying movers in SIPP-EHC will rest primarily on respondent reports, and respondents may be asked to recall address information for a previous residence. The latter is a particularly significant change, as it represents a shift from getting *prospective* address information to *retrospective* address information. Both of these changes increase the likelihood of respondent error in reporting residence history information.

In this paper we evaluate the frequency of reported residences and the accuracy of reported address information in the SIPP-EHC field tests compared to the 2008 production SIPP. We assess the accuracy of the detailed address information in SIPP-EHC by matching the reported addresses to a national address database, and we assess the frequency of reported residences by comparing the number of moves reported in SIPP-EHC to a matched time period and geography sample from the 2008 production SIPP.

The first section of our paper describes production SIPP and SIPP-EHC and how each survey collects residence history information. In the second section, we focus on the quality of reported address data in the 2011 SIPP-EHC field test. Next, we compare mover rates in the two SIPP-EHC field tests to the 2008 SIPP sample. The final section briefly discusses future research.

Residence History Information in Production SIPP

The production SIPP is a longitudinal survey that is nationally representative of the noninstitutionalized U.S. population, with an initial sample size of about 50,000 households. SIPP panels generally last between two and four years, with interviews conducted every four months (called waves). Production SIPP sample members are divided into four roughly equal groups, called rotation groups. Each month, one rotation group is interviewed about the previous four-month reference period, meaning that the calendar months of the reference period will not completely overlap for any two rotation groups within a given production SIPP wave.[1] Interviews include both core questions, asked in each wave, and topical module items, asked only in specific waves.

Residence history data in production SIPP are collected both in the Migration History topical module, asked during the second wave of the 2008 panel, and as part of the core data collected every wave. The Migration History topical module collects data on the year and month in which a respondent moved into his or her current residence, the month and year moved into the previous residence, the tenure status of both current and previous residences, and information on place of birth and nativity status. This topical module is asked only during the second wave of a panel and is limited to respondents who are age 15 and over.

Each wave production SIPP follows original sample members 15 years of age or older when they change addresses, as well as children who continue to live with an original sample member. If respondents move, field representatives rely on several methods to locate them. After the initial interview with a household, field representatives ask for the name and contact information of someone who can provide the new address in the event that an entire household moves. Field representatives may also contact neighbors, employers, mail carriers, real estate companies, rental agents, or postal supervisors to locate original sample members who have moved. In the event that a respondent has moved over 100 miles away from his original residence, the field representative will attempt to contact the respondent by phone. Respondents who cannot be located are dropped from the survey.

Production SIPP only collects data on one move per person per four-month reference period. If a respondent has moved multiple times during the reference period, address information is only collected for the most recent move. This reduces the burden on respondents to report detailed address information about previous residences, but it also may lead to an underestimation of the true mover rate across a production SIPP panel. Residence history information is contained in two variables: (1) an address ID variable, which indicates the month a respondent changes residences, and (2) a mover flag, which indicates whether the respondent moved in a given month and, if so, the distance of the move (intracounty, intercounty same state, interstate).

---

[1] Appendix table A.1 graphically displays the rotation groups, waves, and reference months for the 2008 SIPP panel.

Residence History Information in SIPP-EHC

When fully implemented in 2014, the SIPP-EHC will be a nationally representative sample of the U.S. noninstitutionalized population with four waves of data, spanning four calendar years. The 2010 and 2011 SIPP-EHC field tests consist of smaller samples drawn from self-representing primary sampling units (PSUs) and high-poverty geographies within the production SIPP sampling frame. The 2010 field test initial sample included 7,982 households, of which 5,185 yielded complete interviews. The sample consisted mainly of large cities in California, New York, and Texas, and, not surprisingly, 41.0 percent of the total sample lived in these three states. The sample predominantly consists of younger individuals (median age of 30), renters (67.0 percent), and Hispanics (51.0 percent), further suggesting that the sample households are not representative of the U.S. population.[2] The 2011 field test initial sample included 4,055 households, of which 2,596 yielded complete interviews. The 2011 sample, while slightly more geographically diverse than the 2010 sample, is again not nationally representative with a median age of 30, 64.5 percent renter, and 44 percent Hispanic.[3]

In SIPP-EHC, the event history calendar (EHC) is located in the middle of the instrument, after the demographics section. The EHC collects data on residence history, marital history, education, employment, program participation, and health insurance. To aid respondent recall, the calendar is structured to visually display timelines on the same screen for each of these domains. The calendar for SIPP-EHC has a one-year reference period with months as the unit of analysis. For each domain, respondents report beginning and end month spells. These spells are then visually displayed in the EHC, allowing the field representative to easily identify events that coincide, such as the start of a new job and a change of residence and to use these events to help respondents remember subsequent events. In addition to visually displaying spell lengths, the survey also asks about landmark events that may have occurred during the year (such as a birth, a marriage, or anything else that may stand out as a significant event in the respondent's life).

Within each domain spell, the instrument asks additional questions of each respondent. For example, full address information is collected on each residence where the respondent lived at least one month during the reference year, along with type of living quarters, tenure status (own/rent), receipt of rent subsidy, receipt of housing voucher, and the respondent reported reason for moving to the residence.[4] Up to five addresses may be reported in the EHC. Since many residence spells will be left-censored by the one-year reference period, respondents are also asked for the year and month they moved to the address that they lived in during the first month (January) of the reference year.

There are a number of advantages to the collection of residence history in the SIPP-EHC. First, the universe is broader. In production SIPP the Migration History topical module only goes to respondents who are at least 15 years of age, in SIPP-EHC all respondents get the questions. Additionally, respondents are asked to report up to five addresses in the EHC, rather than the production SIPP method of only allowing one move per four-month reference period. In production SIPP, a large number of respondents have left the sample before the Wave 2 Migration History topical module. This sample attrition combined with production SIPP's current method of using rotation groups, makes calculating yearly or seasonal migration rates difficult. With a year-long reference period, calculating annual and seasonal migration rates will be significantly easier in SIPP-EHC.

While there are clear advantages to collecting residency data in the EHC, this shift in data collection method raises a number of questions about the accuracy of the residence history information collected in SIPP-EHC. Collecting residence information will be more dependent on receiving accurate address information from respondents. The current production SIPP method of only allowing one move per reference period ensures that address information reported by respondents will likely be for a current address. In SIPP-EHC respondents will be asked to report detailed address information about previous residences. Additionally, with a yearlong reference period, it is probable

---

[2] The same numbers for the January, 2009 full SIPP sample are 31.4 percent renter, 15.7 percent Hispanic, and a median age of 36.

[3] The same numbers for the January, 2010 full SIPP sample are 32.1 percent renter, 16.0 percent Hispanic, and a median age of 36.

[4] The respondent is asked to report all residence spells during the reference year that lasted at least the majority of one month. These residence spells can include time in group quarters, medical institutions, emergency or transitional shelters, unoccupied tent or trailer sites, and cars or vans.

that some respondents may be asked to provide proxy address information for people with whom they did not live with for the entire reference period.

Research Goals

Based on the differences between the collection of residence history data in production SIPP and SIPP-EHC, we have identified two research questions: (1) When respondents in SIPP-EHC report detailed previous residence information, do they provide usable information? (2) Are there differences in the number of reported moves in production SIPP and the SIPP-EHC? To test the first research question, we match the addresses collected in the 2011 SIPP-EHC field test to a national address database. We report the percentage of addresses successfully matched at the block, place, county, and state level. To test the second research question, we compare the monthly mover rates reported in the 2010 SIPP-EHC (covering calendar year 2009), the 2011 SIPP-EHC (covering calendar year 2010), and the 2008 production SIPP (covering calendar years 2009 and 2010). We primarily focus on whether there are time trends in the likelihood of respondents to report moving in the two SIPP-EHC field tests, indicating memory decay or misreporting. We conclude by describing the 2012 SIPP-EHC field test and discussing future research projects.

Quality of Address Information in SIPP-EHC

To evaluate the quality of the detailed address information provided by respondents in SIPP-EHC, we use ArcGIS, a powerful geographic information software package, to spatially match the respondents' address information to street addresses and Census boundary information. The boundary information comes from the 2010 Census Topologically Integrated Geographic Encoding and Referencing (TIGER)/Line files, a database that includes spatial extracts from the Census Bureau's Master Address File/TIGER database, containing features such as roads, railroads, rivers, as well as legal and statistical geographic areas. These shapefiles are designed specifically for use with geographic information software such as ArcGIS.

We decided to use ArcGIS mainly because of time constraints. Normally, addresses from SIPP are geocoded down to the block level using the Census Bureau's proprietary autocoding software, which matches addresses using the same TIGER/Line files. Eventually, this software will also be used to geocode addresses from the production SIPP-EHC.

Geocoding software includes tolerance settings that allow for user control over sensitivity to spelling and number agreement in reported addresses. Matches do not have to be exact because the coding algorithm in this software allows for equivocations, such as using sound-alike values of letters (for example, m=n, f=ph, etc.) and reversing consecutive letter combinations (ie=ei). Each equivocation is assigned a numeric value representing a confidence level in the match, with exact matches receiving the best score (usually 100). The responses have to match reference file entries (i.e. TIGER/Line files) with a relatively high level of confidence in order for an automated computer match to be accepted. Geocoding software further provides the user with the ability to match addresses manually that do not fit the criteria for an automated computer match. This process, called clerical coding, is standard practice for large Census Bureau Surveys like SIPP, and the American Community Survey (ACS). Because of differences in tolerances settings and coding algorithm between ArcGIS and the Census autocoder, our results may not fully approximate the results from the Census autocoder. Our intent is to eventually compare SIPP-EHC address match rates from ArcGIS to those of the Census autocoder to evaluate the quality of geocoded addresses across the two systems.

Table 1 shows the match rates for addresses from the 2011 SIPP-EHC, both overall and for SCIF (Survey Control Input File) and non-SCIF addresses.[5] SCIF addresses are original sample addresses, while non-SCIF addresses are respondent-provided addresses for previous residences. We created an address file by pulling all unique addresses provided by all households in the 2011 SIPP-EHC who had completed interviews. We removed from the file any address whose information was totally blank; we included any address that had at least one piece of address information present, even if it was only the country (the instrument does not ask for detailed address information for

---

[5] The SCIF is the file with the original interview address for each case.

any address outside of the United States and Puerto Rico). [6] After these deletions, our working address file contained 3,295 addresses, of which we were able to match about 91 percent (3,008) to a street address in the TIGER/Line reference files. About 79 percent of addresses (2,589) were exact matches with no corrections needed. Another 11 percent were matchable after making minor corrections, such as adding street directional designations (e.g., W, E) or correcting street suffixes (e.g., substituting "Drive" for "Road"). For the remaining 1.7 percent of matched addresses, we were able to identify them using external data such as the United States Postal Service webpage or Google Maps.

We were not able to match 8.7 percent (287) of addresses to the street level. However, this does not mean that we do not know anything about them: Some addresses are matched to the street with no street number (2.1 percent, 70), while others matched at the city/county level (3.8 percent, 126). Additionally, 35 addresses (1.1 percent) are outside of the U.S., so the instrument did not ask for the detailed information for those instances.[7] Only 1.4 percent of addresses (46) contain unusable data with incorrect street names, and 0.3 percent (10) only include state or country codes. The matches are determined both by automatic or computer coding (89.5 percent) and by computer-assisted clerical coding (10.5 percent).[8]

The majority of the addresses are derived from the sampling frame (2,579) and not reported by respondents (SCIF).[9] As expected, the match rate for original sample addresses (99.9 percent) is extremely high. Of these addresses, 93.0 percent are exact matches, 5.3 percent are matched through minor corrections, and 1.6 percent through investigation. Of the two unmatched addresses, one has an invalid street address, the other only a post office box.

The respondent provided addresses (non-SCIF) in the final column of Table 1 represent the retrospective address information collected in the EHC. Of these 716 addresses, we were able to successfully match 431 (60.2 percent) of them. However, only about one quarter (26.5 percent) are exact matches, while 31.4 percent matched with minor corrections and 2.2 percent matched after more detailed investigation. Even though nearly 40.0 percent of these addresses cannot be matched at the street level, most of them can be matched to the street with no street number (9.8 percent) or at the city/county level (17.5 percent). An additional 4.9 percent are out of the country. Addresses with unusable data include 6.3 percent with incorrect street names and 1.4 percent with only a state or country code.

Table 2 includes the Census Bureau's block-level autocoding results for movers from waves 2 through 8 of the 2008 production SIPP. These data are presented for comparison purposes and are most comparable to the respondent-provided address information in Table 1. However, we recommend caution when interpreting these results. First, the results are produced using the Census Bureau's autocoder rather than ArcGIS, so apparent differences in the results may be due to the method of geocoding used. Second, because production SIPP only allows one move per reference period, these addresses are for current residences at the time of interview, rather than previous addresses as reported in SIPP-EHC, so we would expect them to be of generally higher quality. Finally, the data in Table 1 are coded to the street level rather than the block level, and data unmatched at the street level are matched to street with no number, city, and county, but not block.

Overall, the percentage of addresses successfully coded at the block level range between 89.5 percent and 91.7 percent across waves 2 and 8.[10] Many of these addresses are matched using automated (computer) coding (78.7-84.2 percent), with the remaining matched using clerical coding. The autocoding results from production SIPP appear to yield a greater proportion successfully coded addresses. However, this is somewhat expected given SIPP-EHC's greater reliance on respondent-reported previous addresses. Overall, about 87.5 percent of the respondent reported

---

[6] To protect respondent confidentiality and to prevent any inadvertent disclosure of information, before geocoding the addresses we removed all personally identifying information that might tie the addresses back to specific respondents. Thus, our file consisted only of a list of fully anonymized addresses.

[7] This number is higher than the 0.4 percent of respondents in the 2009 ACS that reported living in a foreign country on a question about their residence one year ago.

[8] Automatic coding includes addresses that were not matched and could not be subsequently matched through clerical coding.

[9] 17.6 percent of respondents in the 2011-SIPP EHC reported moving at least one time during the reference year.

[10] The wavely totals in Table 2 include only addresses with at least two digits of address information, such as a state code.

addresses in Table 1 can be coded to at least the county level (matched to street, matched to street with no number, matched to city, and matched to county). This suggests that respondents in SIPP-EHC generally reported usable address information for previous residences.

Table 3 further breaks down the coding results from Table 1 by state. 801 of the 3,295 addresses are from California. New York is second with 454 addresses, Texas third with 376, and Florida fourth with 289. Focusing on total addresses, Georgia (85.4 percent) has the lowest overall match rate, while New Jersey (98.2 percent) has the highest rate. However, the overall match rates are heavily influenced by the percentage of respondent reported (non-SCIF) addresses. For example, only 60.4 percent of the addresses from Arizona are original sample addresses, the lowest of any state, compared to 88.0 percent in Maryland. Focusing only on the four states with the most respondent-reported addresses – California, New York, Texas, and Florida – there is some variation in the percentage of respondent-reported addresses that we successfully matched. About half of respondent-reported addresses in California (48.7 percent) are unmatched, compared to fewer than 40 percent in New York (34.4 percent), Texas (31.8 percent), and Florida (32.6 percent).

These results show that the quality of the reported address data is of high quality in both the production SIPP and the SIPP-EHC. It is clear that the SIPP-EHC data leads to fewer automatic matches, but that is not unexpected given the retrospective nature of the survey and its reliance on the respondent to recall very specific pieces of information.[11]

Mover Rates in SIPP and SIPP-EHC

In this section we compare mover rates in the 2010 and 2011 SIPP-EHC test data to those from the 2008 production SIPP. Because of differences in how movers are tracked across the two surveys, sample attrition, and sample limitations in SIPP-EHC, we did a number of things to make the two data sources as comparable as possible. First, we matched the data to time period. Specifically, we compare the 2010 SIPP-EHC mover rates (covering the 2009 calendar year) to a 2009 calendar year file from the Waves 2-5 of the 2008 production SIPP. Likewise, we compare the 2011 SIPP-EHC data to a 2010 calendar year file from Waves 5-8. Second, we present the production SIPP mover rates for both the overall sample and a sample matched to the SIPP-EHC sample geography to account for the lack of geographic diversity in SIPP-EHC.[12] Finally, only moves in which the respondent remained in sample are included in the production SIPP mover estimates. This is because it is not always possible to identify the month of move for respondents who move and cannot subsequently be located in the production SIPP sample, making comparisons to SIPP-EHC difficult. Consequently, our production SIPP mover estimates do not represent the total number of movers in production SIPP, but the number of movers that remain in universe and are successfully located by field representatives.

Preliminary weights for the 2010 and 2011 SIPP-EHC were created by the Demographic Surveys Division. The SIPP-EHC weights adjust the sample to correspond to the weighted 2008 production SIPP matched geography samples. More specifically, the 2010 SIPP-EHC weights are based on the January, 2009 production SIPP matched sample population. The 2011 SIPP-EHC weights are based on the January, 2010 production SIPP matched sample population. For example, the weighted population for the geography matched production SIPP sample for January, 2010 is 29,931,714 and the weighted population total for the 2011 SIPP-EHC sample from January is 29,832,882. However, the weighting does not completely mitigate the effects of sample attrition in the production SIPP. Because we use the January weight, our production SIPP mover estimates for later calendar months are subject to the effect of sample attrition. The two SIPP-EHC field tests are essentially Wave 1 interviews, so there is no sample attrition

---

[11] Another consideration is the use of edited data. All of the results presented from SIPP-EHC come from unedited data. However, we have written edits for the residence history section that should improve the quality of the data and therefore lead to an even higher match rate. For instance, if one person in a household provides an incomplete address, but another household member provides a full address (and the two lived together during the relevant period), then we will replace the incomplete information with the complete information, allowing us to match an address that was previously unmatchable.

[12] In analyses not shown, we compared the two SIPP-EHC samples to the matched time period and geography SIPP samples on tenure, race, Hispanic origin, language (speak non-English in the home) and age and found the matched SIPP samples to be highly comparable to the SIPP-EHC field tests. For example, 66.7 percent of respondents in the SIPP 2010 matched sample are renters, compared to 67.0 percent in the 2010 SIPP-EHC.

across the one-year reference period. For production SIPP, there is significant sample loss across the waves needed to construct the 2009 and 2010 calendar year files. For example, by Wave 8, 38 percent of original sample households have left the production SIPP sample. Furthermore, by Wave 8, about 4.5 percent of original sample households are lost because they cannot be located after a move.[13]

To partially account for sample attrition, we calculate mover estimates for the entire calendar year and by calendar month. The monthly mover rates are calculated based on the number of respondents in sample in the month of interest, somewhat reducing the impact of sample attrition in production SIPP. Despite the impact of sample loss on comparability across the two surveys, this issue also highlights the strength of SIPP-EHC data. One calendar year of data is collected in SIPP-EHC before normal sample attrition that inevitably occurs in follow-up interviews.

Table 4 describes the overall mover rates for the weighted 2010 SIPP-EHC and matched production SIPP samples. We present production SIPP numbers for both the full sample and 2010 SIPP-EHC geography matched sample. The mover estimates from the production SIPP samples are for the 2009 calendar year, matching the reference period of the 2010 SIPP-EHC. Again, please interpret these numbers with some caution due to sample attrition across the production SIPP surveys.[14] The percentage of the 2010 SIPP-EHC sample that reported moving in 2009 is 14.7 percent.[15] This number is substantially greater than the 9.2 percent of respondents in the full production SIPP sample, and 10.3 percent in the production SIPP matched geography sample, who moved and were subsequently successfully located. Respondents in the 2010 SIPP-EHC also include a greater number of repeat movers than in the 2008 production SIPP. About 1.7 percent of all respondents in the 2010 SIPP-EHC report multiple moves during 2009, compared to only 0.9 percent of respondents in production SIPP and 0.7 percent in the matched production SIPP sample.

In Table 5, we present the results for the 2011 SIPP-EHC compared to 2008 production SIPP. 17.6 percent of respondents in the 2011 SIPP-EHC report moving in 2010, quite a bit higher than the 8.8 percent of respondents in the full sample and 10.9 percent in the matched production SIPP sample who moved and were successfully located. Similar to the results from the 2010-SIPP EHC, the number of respondents who report multiple moves in the 2011 SIPP-EHC (2.1 percent) is greater than the number of multiple movers identified in the 2010 calendar year of the 2008 production SIPP (1.2 percent).

Figure 1 shows mover trends by month for calendar year 2009. The figure shows three lines: (1) the 2008 SIPP, (2) the matched geographic subset from the 2008 SIPP, and (3) the 2010 SIPP-EHC.[16] Given the results from the previous section, we expect the monthly mover proportions to be greater in SIPP-EHC than in production SIPP. Additionally, differences between the surveys should be fairly uniform across months, regardless of seasonal move

---

[13] For the 2009 and 2010 SIPP calendar year files, we investigated the extent to which sample attrition altered the composition of the full and matched samples. We found that, for both calendar years, and both the full and matched samples, the sample composition at the end of the year tended to be slightly older and more likely to be homeowners compared to earlier months. This suggests that the SIPP samples lose younger respondents and/or renters, two groups that move frequently.

[14] The denominator for the 2008 SIPP mover estimates includes all respondents who were in sample for at least one month during the given calendar year. Because of sample attrition, this estimate inevitably underestimates the true mover rate across the sample. For example, 101,118 respondents in SIPP were in universe for at least one month between February and December of 2009. However, the most respondents in universe for a given calendar month during this time was 97,045 in February, and the lowest was 90,393 in December.

[15] The estimates in this report (which may be shown in text, figures, or tables) are based on responses from a sample of the population and may differ from the actual values because of sampling variability or other factors. As a result, apparent differences between the estimates for two or more groups may not be statistically significant. All comparative statements have undergone statistical testing and are significant at the 90 percent confidence level unless otherwise noted. For further information on statistical standards and the computation and use of standard errors, go to http://www.census.gov/sipp/source.html or contact Tracy Mattingly of the Census Bureau's Demographic Statistical Methods Division on the internet at Tracy.l.Mattingly@census.gov.

[16] A tabular version of this figure is included as Appendix Table A.3.

patterns.[17] The results suggest the differences between the proportions of movers by month between the SIPP-EHC sample and the 2009 calendar year production SIPP matched file are greater at later calendar months. For example, the percent of the 2010 SIPP-EHC sample that reported moving in the month of February (1.0 percent) is not statistically different from the 1.2 percent of the 2009 production SIPP calendar year matched geography sample that reported moving in February. However, by December, 1.8 percent of the 2010 SIPP-EHC sample reports a move, compared to 0.9 percent of the production SIPP 2009 calendar year matched sample. Moreover, there is limited evidence of a seasonal move trend in the 2010 SIPP-EHC data. While the reported number of moves appears to increase in the summer months, the proportion of movers remains high in November and December, where the monthly proportion of movers declines in the two production SIPP samples in these months. This pattern of high mobility in November, and December in the 2010 SIPP-EHC and relatively low mobility in earlier months such as February and March does not correspond to expected seasonal move patterns and may indicate memory decay or respondent fatigue.

Figure 2 is arranged identically to Figure 1 and presents mover trends by month for calendar year 2010.[18] Like the pattern observed in the 2010 SIPP-EHC, later months in the 2011 SIPP-EHC reference year (i.e. October and November) have greater numbers of movers than the matched production SIPP sample, while the number of movers in the earlier months (February/March) is not statistically different from the matched production SIPP sample. However, the two production SIPP samples show fairly low percentages of movers, even in summer months, suggesting sample attrition may be distorting the production SIPP mover trends. There is generally more evidence of the seasonality of moves in the 2011 SIPP-EHC data, than in the 2010 data. For example, the proportion of the sample that reported a move decreased from 2.4 percent in October to 1.6 percent in December.

Overall, the two SIPP-EHC field tests yielded more movers than the production SIPP samples. While this may be largely because of our decision to omit movers in production SIPP who left the sample, our findings suggest that the amount of usable mover data appears to be greater in SIPP-EHC than in production SIPP. However, there is also some evidence that respondents in SIPP-EHC report fewer moves in earlier calendar months than in later months. This could indicate recall problems or respondent fatigue with reporting detailed address information.

Future Research

We plan to undertake a number of future research projects to further evaluate the quality of address information collected in SIPP-EHC.

(1) The residence history information in SIPP-EHC can be used to create a distance moved recode variable indicating whether a reported move was intracounty, intercounty within the same state, or interstate. We plan to compare the actual distance moved, calculated using the geocoded addresses, to this recode variable. We are primarily interested in identifying short-distance interstate moves, and long-distance intracounty moves and how frequently these situations occur.

(2) In addition to collecting detailed residence history information, SIPP-EHC also collects detailed address information for up to seven employers for each adult respondent, information not collected in production SIPP. In future research, we plan to evaluate the quality of the employer address information using the geocoding methods in this paper.

(3) Because of time constraints, the addresses in this paper were geocoded using ArcGIS, rather than the Census auto-coder. We plan to run the SIPP-EHC tests data addresses through the Census auto-coder and compare the results to those from ArcGIS. This will allow us to assess the effectiveness of both geocoding platforms.

(4) We plan to look at beginning and end months of spells for other topics in the SIPP-EHC to see if the pattern of more spells at later months of the reference year is unique to the residence history section, or more general to the EHC.

---

[17] Mover rates tend to be highest during the summer months and lowest during winter. See Marlay, Matthew C., and Alison K. Fields (2010). *Seasonality of Moves and the Duration and Tenure of Residence: 2004.* Current Population Reports, P70-122. U.S. Census Bureau, Washington, DC.

[18] A tabular version of this figure is included as Appendix table A.4

The Census is currently preparing a 2012-SIPP-EHC field test. This version of the survey will be a Wave 2 interview, following respondents from the 2011 SIPP-EHC field test for a second year. The 2012 field test will provide an opportunity to study the impact of sample attrition on Wave 2 interviews. One of the strengths of SIPP-EHC is that the first wave of the survey will collect an entire year's worth of information, rather than the four months of data collected in the first wave of production SIPP. This is important because sample attrition is greatest in production SIPP after the first wave, meaning a large number of respondents exit the survey only having provided four months of information and without answering a number of the topical modules. However, the yearlong gap between waves in SIPP-EHC means successfully identifying movers between waves is particularly important in SIPP-EHC. The Census Bureau is planning to conduct a locating experiment on households that will be part of the 2012 SIPP-EHC field test. The households will be split into three groups. The first group will receive a self-enclosable address change postcard indicating whether the whole household has moved, part of the household has moved, the household intends to move, and updated contact information. This group will also receive a small monetary incentive for responding to the address update. The second group will receive the address update postcard, but no incentive. The third group will be the control, and will not receive the address postcard. This experiment is intended to help the Bureau develop more effective methods for locating movers.

Conclusion

The purpose of SIPP-EHC is to reduce both the survey cost and respondent burden of collecting 4 years worth of longitudinal data on employment and program participation, while still providing data of similar content and structure to the current production SIPP. The shift to a year-long reference period in SIPP-EHC increases the reliance on respondents to recall information, particularly for the residence history section, where respondent in SIPP are never asked to recall retrospective address information.

In this paper we evaluated the residence history data collected in SIPP-EHC by focusing on the quality and frequency of reported residence information compared to the 2008 production SIPP. Our results are cautiously optimistic. A majority of the detailed retrospective address information reported in the 2011 SIPP-EHC was successfully coded to at least the county-level, with almost 60 percent coded to the street-level. While these numbers are lower than the block-level coding success rates for the 2008 production SIPP, this was expected, because current SIPP does not collect retrospective address information. The comparisons of 2010 and 2011 SIPP-EHC mover rates to those of calendar year files from the 2008 production SIPP suggest that SIPP-EHC collects more usable mover data than production SIPP, as confirmed by higher overall mover rates. Evidence of seasonal move patterns in the 2011 SIPP-EHC was observed, with more movers in the summer months than winter, although this pattern was not present in the 2010 SIPP-EHC. There was, however, some evidence of potential memory decay, or respondent misreporting of residences during the early months of the EHC in both tests surveys.

Table 1: Geocoding Results from 2011 SIPP-EHC

	Total		From SCIF			
			Yes		No	
	Number	Percent	Number	Percent	Number	Percent
Address Status	3,295	100.0	2,579	100.0	716	100.0
Able to Match	3,008	91.3	2,577	99.9	431	60.2
Exact	2,589	78.6	2,399	93.0	190	26.5
Minor corrections	362	11.0	137	5.3	225	31.4
After investigation	57	1.7	41	1.6	16	2.2
Unable to Match	287	8.7	2	0.1	285	39.8
Invalid street address	46	1.4	1	0.0	45	6.3
Street but no number	70	2.1	0	0.0	70	9.8
City	116	3.5	1	0.0	115	16.1
County	10	0.3	0	0.0	10	1.4
State	7	0.2	0	0.0	7	1.0
Country	3	0.1	0	0.0	3	0.4
Non-U.S. address	35	1.1	0	0.0	35	4.9
Match Type	3,295	100.0	2,579	100.0	716	100.0
Automatic	2,950	89.5	2,386	92.5	564	78.8
Manual	267	8.1	128	5.0	139	19.4
Picked from map	78	2.4	65	2.5	13	1.8

Table 2: Autocoder Results from 2008 Production SIPP (Block-level)

| Wave | Total | | Matched | | | | | |
| | Number | Percent | Total | | Auto | | Clerical | |
			Number	Percent	Number	Percent	Number	Percent
Wave 2	1,785	100.0	1,597	89.5	1,428	80.0	169	9.5
Wave 3	2,172	100.0	1,962	90.3	1,741	80.2	221	10.2
Wave 4	2,613	100.0	2,418	92.5	2,154	82.4	264	10.1
Wave 5	2,181	100.0	1,974	90.5	1,741	79.8	233	10.7
Wave 6	2,282	100.0	2,089	91.5	1,814	79.5	275	12.1
Wave 7	2,195	100.0	1,987	90.5	1,727	78.7	260	11.8
Wave 8	1,796	100.0	1,646	91.7	1,512	84.2	134	7.5

Table 3: Geocoding Results from 2011 SIPP-EHC by State

	From SCIF														
	Overall					Yes					No				
		Matched		Unmatched			Matched		Unmatched			Matched		Unmatched	
	Total	Number	Percent	Number	Percent	Total	Number	Percent	Number	Percent	Total	Number	Percent	Number	Percent
All Addresses	3,295	3,008	91.3	287	8.7	2,579	2,577	78.2	2	0.1	716	431	13.1	285	8.6
Arizona	111	95	85.6	16	14.4	67	67	60.4	0	0.0	44	28	25.2	16	14.4
California	801	725	90.5	76	9.5	645	645	80.5	0	0.0	156	80	10.0	76	9.5
Connecticut	51	47	92.2	4	7.8	41	41	80.4	0	0.0	10	6	11.8	4	7.8
Florida	289	270	93.4	19	6.6	231	231	79.9	0	0.0	58	39	13.5	19	6.6
Georgia	41	35	85.4	6	14.6	30	30	73.2	0	0.0	11	5	12.2	6	14.6
Illinois	138	131	94.9	7	5.1	113	113	81.9	0	0.0	25	18	13.0	7	5.1
Indiana	53	48	90.6	5	9.4	37	37	69.8	0	0.0	16	11	20.8	5	9.4
Louisiana	67	62	92.5	5	7.5	53	53	79.1	0	0.0	14	9	13.4	5	7.5
Maryland	50	49	98.0	1	2.0	44	44	88.0	0	0.0	6	5	10.0	1	2.0
Massachusetts	115	108	93.9	7	6.1	88	88	76.5	0	0.0	27	20	17.4	7	6.1
Michigan	83	78	94.0	5	6.0	64	64	77.1	0	0.0	19	14	16.9	5	6.0
Missouri	63	60	95.2	3	4.8	49	49	77.8	0	0.0	14	11	17.5	3	4.8
New Jersey	57	56	98.2	1	1.8	48	48	84.2	0	0.0	9	8	14.0	1	1.8
New York	454	432	95.2	22	4.8	390	390	85.9	0	0.0	64	42	9.3	22	4.8
Ohio	141	132	93.6	9	6.4	110	110	78.0	0	0.0	31	22	15.6	9	6.4
Oklahoma	63	57	90.5	6	9.5	48	47	74.6	1	1.6	15	10	15.9	5	7.9
Pennsylvania	160	154	96.3	6	3.8	139	139	86.9	0	0.0	21	15	9.4	6	3.8
Tennessee	68	63	92.6	5	7.4	49	49	72.1	0	0.0	19	14	20.6	5	7.4
Texas	376	348	92.6	28	7.4	291	290	77.1	1	0.3	85	58	15.4	27	7.2
Virginia	51	49	96.1	2	3.9	42	42	82.4	0	0.0	9	7	13.7	2	3.9
Other States	25	9	36.0	16	64.0	0	0	0.0	0	0.0	25	9	36.0	16	64.0

Table 4: 2010 SIPP-EHC Matched to 2008 SIPP, 2009 Calendar Year (Waves 2-5)

Variable	2009 Calendar Year, Waves 2-5 of 2008 SIPP			2009 Calendar Year, Matched to 2010 Geography			2010 Unedited SIPP-EHC		
	Number	Percent	C.I.	Number	Percent	C.I.	Number	Percent	C.I.
Number of Moves[1,2]	107,066	100.0	-	7,777	100.0	-	14,726	100.0	-
0	97,695	90.9	0.2	6,993	89.7	0.8	12,568	85.3	0.5
1	8,518	8.3	0.2	734	9.6	0.7	1,910	13.0	0.5
2 or more	853	0.9	0.1	50	0.7	0.2	248	1.7	0.2

1. Because of an instrument error that lead to erroneous respondent reports of moving in January, the estimates do not include January moves.
2. In this table the reported sample sizes are unweighted, but the proportions are weighted to January, 2009 of the production SIPP matched geography sample.

Table 5: 2011 SIPP-EHC Matched to 2008 SIPP, Wave 5-8 2010 Calendar Year

Variable	2010 Calendar Year, Waves 5-8 of 2008 SIPP			2010 Calendar Year, Matched to 2011 Geography			2011 Unedited SIPP-EHC		
	Number	Percent	C.I.	Number	Percent	C.I.	Number	Percent	C.I.
Number of Moves[1,2]	101,118	100.0	-	12,282	100.0	-	7,003	100.0	-
0	92,402	91.1	0.2	10,994	89.1	0.1	5,769	82.4	0.7
1	7,912	8.0	0.2	1,153	9.7	0.7	1,087	15.5	0.7
2 or more	804	0.8	0.2	135	1.2	0.6	147	2.1	0.3

1. Because of an instrument error that lead to erroneous respondent reports of moving in January, the estimates do not include January moves.
2. In this table the reported sample sizes are unweighted, but the proportions are weighted to January, 2010 of the production SIPP matched geography sample.
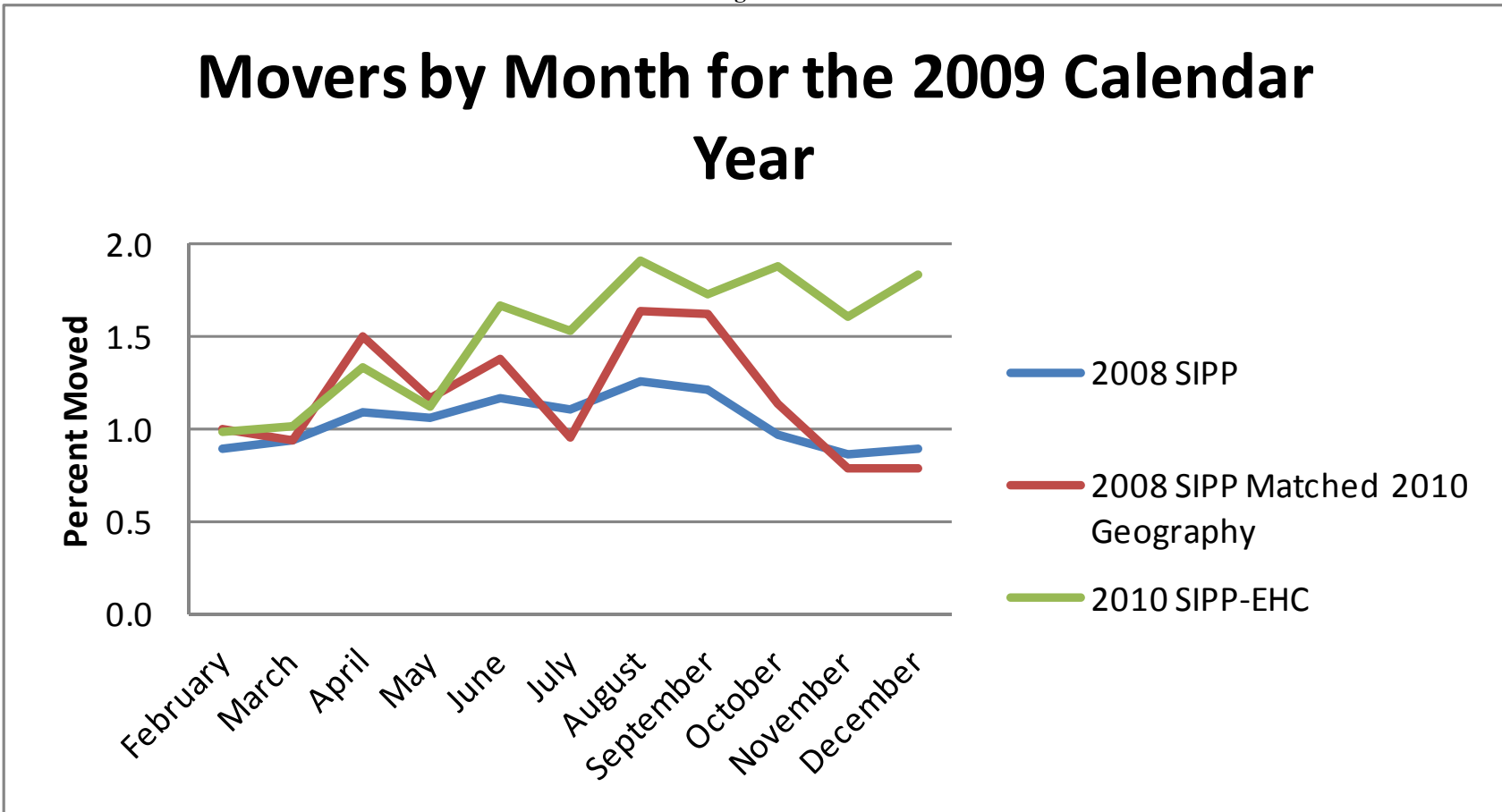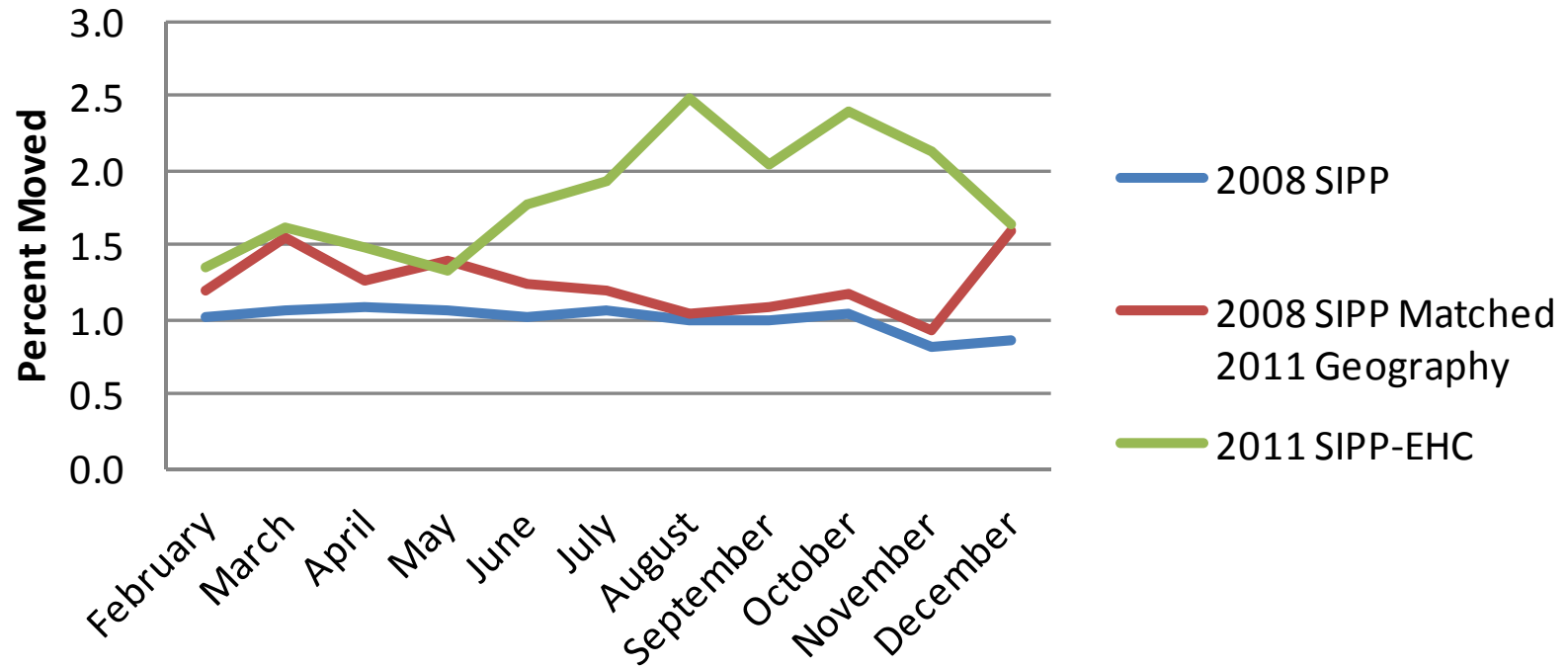
Figure 1

Movers by Month for the 2009 Calendar Year

Figure 2

Movers by Month for the 2010 Calendar Year

Appendix

Table A.1: Rotation Groups, Waves, and Reference Months for the 2008 SIPP Panel

2008 SIPP	2008								2009											
	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Rotation 1	1-1	1-2	1-3	1-4	2-1 (I)	2-2	2-3	2-4	3-1 (I)	3-2	3-3	3-4	4-1(I)	4-2	4-3	4-4	5-1(I)	5-2	5-3	5-4
Rotation 2		1-1	1-2	1-3	1-4	2-1 (I)	2-2	2-3	2-4	3-1 (I)	3-2	3-3	3-4	4-1(I)	4-2	4-3	4-4	5-1(I)	5-2	5-3
Rotation 3			1-1	1-2	1-3	1-4	2-1 (I)	2-2	2-3	2-4	3-1 (I)	3-2	3-3	3-4	4-1(I)	4-2	4-3	4-4	5-1(I)	5-2
Rotation 4				1-1	1-2	1-3	1-4	2-1 (I)	2-2	2-3	2-4	3-1 (I)	3-2	3-3	3-4	4-1(I)	4-2	4-3	4-4	5-1(I)

2008 SIPP	2010											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Rotation 1	6-1(I)	6-2	6-3	6-4	7-1 (I)	7-2	7-3	7-4	8-1 (I)	8-2	8-3	8-4
Rotation 2	5-4	6-1	6-2	6-3	6-4	7-1 (I)	7-2	7-3	7-4	8-1 (I)	8-2	8-3
Rotation 3	5-3	5-4	6-1	6-2	6-3	6-4	7-1 (I)	7-2	7-3	7-4	8-1 (I)	8-2
Rotation 4	5-2	5-3	5-4	6-1	6-2	6-3	6-4	7-1 (I)	7-2	7-3	7-4	8-1 (I)

Note: In the number sequences, the first indicates the wave and the second indicates the reference month. (I) indicates the interview month. The reference period is the four months preceding the interview month.
Source: U.S. Census Bureau, Survey of Income and Program Participation (SIPP) 2008 panel.

Table A.2: Geocoding Results from 2011 SIPP-EHC by Match Type

	Match Type							
	Total		Automatic		Manual		Picked from Map	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Address Status	3,295	100.0	2,950	100.0	267	100.0	78	100.0
Able to Match	3,008	91.3	2,666	90.4	264	98.9	78	100.0
Exact	2,589	78.6	2,456	83.3	88	33.0	45	57.7
Minor corrections	362	11.0	210	7.1	145	54.3	7	9.0
After investigation	57	1.7	0	0.0	31	11.6	26	33.3
Unable to Match	287	8.7	284	9.6	3	1.1	0	0.0
Invalid street address	46	1.4	46	1.6	0	0.0	0	0.0
Street but no number	70	2.1	67	2.3	3	1.1	0	0.0
City	116	3.5	116	3.9	0	0.0	0	0.0
County	10	0.3	10	0.3	0	0.0	0	0.0
State	7	0.2	7	0.2	0	0.0	0	0.0
Country	3	0.1	3	0.1	0	0.0	0	0.0
Non-U.S. address	35	1.1	35	1.2	0	0.0	0	0.0

Table A.3: 2010 SIPP-EHC Matched to 2008 SIPP, 2009 Calendar Year (Waves 2-5)

Month of Move	2009 Calendar Year, Waves 2-5 of 2008 SIPP			2009 Calendar Year, Matched to 2010 Geography			2010 Unedited SIPP-EHC		
	Total N	Movers	Percent	Total N	Movers	Percent	Total N	Movers	Percent
January[1]	97,630	843	0.8	6,965	55	0.8	14,542	372[2]	2.6
February	97,045	832	0.9	6,869	80	1.0	14,560	144	1.0
March	96,010	867	0.9	6,794	64	0.9	14,577	150	1.0
April	94,996	979	1.1	6,694	102	1.5	14,595	195	1.3
May	94,169	923	1.0	6,650	70	1.2	14,609	167	1.1
June	93,084	1,067	1.2	6,573	89	1.4	14,627	248	1.7
July	91,972	997	1.1	6,444	63	0.9	14,646	220	1.5
August	91,089	1,120	1.3	6,368	91	1.6	14,662	277	1.9
September	90,541	1,065	1.2	6,313	99	1.6	14,679	248	1.7
October	90,145	846	1.0	6,291	79	1.1	14,698	274	1.9
November	90,351	774	0.9	6,388	42	0.8	14,709	240	1.6
December	90,393	821	0.9	6,433	59	0.8	14,726	269	1.8

1. In this table the reported sample sizes are unweighted, but the proportions are weighted to January, 2009 of the production SIPP matched geography sample.
2. The estimate of January movers is unreliable because of an instrument error.

TABLE A.4: 2011 Re-SIPP Matched to 2008 SIPP, Wave 5-8 2010 Calendar Year

Month of Move	2010 Calendar Year, Waves 5-8 of 2008 SIPP			2010 Calendar Year, Matched to 2011 Geography			2011 Unedited SIPP-EHC		
	Total N	Movers	Percent	Total N	Movers	Percent	Total N	Movers	Percent
January[1]	90,034	914	1.0	10,625	169	1.6	6,989	123[2]	1.8
February	89,745	886	1.0	10,707	133	1.2	6,989	95	1.3
March	88,911	902	1.1	10,632	154	1.5	6,989	113	1.6
April	88,087	897	1.1	10,554	146	1.3	6,990	103	1.5
May	87,622	965	1.1	10,561	131	1.4	6,990	90	1.3
June	86,960	921	1.0	10,493	136	1.2	6,991	124	1.8
July	86,291	955	1.1	10,427	119	1.2	6,992	133	1.9
August	85,376	864	1.0	10,334	118	1.0	6,994	177	2.5
September	84,966	885	1.0	10,295	131	1.1	6,996	143	2.0
October	84,422	876	1.0	10,227	118	1.2	6,996	169	2.4
November	84,411	688	0.8	10,210	95	0.9	7,002	149	2.1
December	84,410	722	0.9	10,309	152	1.6	7,003	114	1.6

1. In this table the reported sample sizes are unweighted, but the proportions are weighted to January, 2010 of the production SIPP matched geography sample.
2. The estimate of January movers is unreliable because of an instrument error.