THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

SIPP Longitudinal Household Estimation for the Proposed Longitudinal Definition

No. 13

Lawrence R. Ernst

U.S. Census Bureau

June 1986

U.S. Department of Commerce U.S. CENSUS BUREAU

THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

SIPP LONGITUDINAL HOUSEHOLD ESTIMATION FOR THE PROPOSED LONGITUDINAL DEFINITION

NO. 8605 13
By
Lawrence R. Ernst
Bureau of the Census

SIPP Working Paper #13

June 1986

Acknowledgements

This paper was prepared by Lawrence R. Ernst, Statistical Research Division, Bureau of the Census.

The author would like to thank David Judkins, Statistical Methods Division; Daniel Kasprzyk, Population Division; and Nash Monsour, Jr., Statistical Research Division, for their assistance in reviewing this paper.

Suggested Citation

Ernst, Lawrence R. "SIPP Longitudinal Household Estimation for the Proposed Longitudinal Definitions," SIPP Working Paper Series No. 8605. Washington, D.C.: Bureau of the Census, 1986.

TABLE OF CONTENTS

1. INTRODUCTION

In [1] and [3] weighting procedures were developed for obtaining unbiased longitudinal household (LHH) and family estimates for SIPP. As noted in these papers and also in [2], except for certain LHH definitions it does not appear possible with the original SIPP operating procedures to develop an unbiased weighting procedure without serious drawbacks. These drawbacks include the assignment of positive weights to some LHH's that were not in sample for their entire period of existence during the life of the panel, or the lack of sufficient information to assign weights to some LHH's because some of the necessary information pertained to time periods when the LHH's were not in sample. These previous papers were handicapped by the fact that the choice among weighting methods and the changes in operational procedures necessary to overcome the problems just described are dependent on the LHH definition adopted for SIPP, and no LHH definition had been agreed upon at the time these papers were written. Now, however, a definition does exist which has been tentatively adopted (Attachment A). Although this particular definition does require changes in SIPP operating procedures in order to obtain unbiased estimates without the drawbacks previously mentioned, only two relatively simple changes are required. One of these changes has already been implemented for this purpose and the other is planned.

In Section 2 of this paper, unbiased weighting will be discussed. This discussion will include details of the necessary operational changes and the reasons for them. Two unbiased

weighting procedures appropriate for use with the modified operational procedures are described and compared. Several other unbiased weight procedures that were previously described in [3] are also discussed.

In Section 3 a set of proposed adjustments to the set of unbiased LHH weights is detailed. All of the adjustment steps typically found in the demographic surveys conducted by the Census Bureau are included, but there are also come important complications that are unique to SIPP LHH estimation.

Basic knowledge of SIPP, including the design of this survey, which can be obtained from [10], is assumed in this paper. Also assumed is a general understanding, which can be obtained from [5], of the various stages of weighting commonly used in the demographic surveys conducted by the Census Bureau.

2. UNBIASED WEIGHTING

2.1 Preliminaries

Some notation and terminology to be used in this paper will be presented here.

The LHH definition referred to in this paper is given in Attachment A.

Let t_B, t_F denote the first and last month respectively for a SIPP panel. This can be taken to be the first and last reference month respectively common to the four rotation groups that comprise the panel.

When reference is made to the period of existence of a LHH, it is understood that this is the intersection of the actual

period of existence, which may be many years, and the interval for which estimates are made. For example, if the actual period of existence of a household is from July 1985 through May 1986, then with respect to estimates for the calendar year 1985 the period of existence of this household is July 1985 through December 1985.

Finally, unless otherwise stated, the first and last month of existence of a LHH are denoted by t_1, t_2 respectively.

2.2 Two Unbiased Weighting Procedures

Presented here are the two unbiased weighting procedures that we consider the most appropriate with the LHH definition in the sense that these weighting procedures require the minimal amount of change in the operating procedures to avoid the problems of the type mentioned in the Introduction.

To describe these weighting procedures, we first associate an unbiased person weight to each "adult" (that is an individual who was at least 15 years of age by t_F) who was in the cross-sectional universe at any time during the life of the panel. For all original sample persons, that is those adults who were in sample for Wave 1 (except those who were in sample erroneously), this weight is the reciprocal of the probability of selection of the individual's Wave 1 household (HH). For all other adults, including those outside the universe during Wave 1, the weight is zero.

The two weighting procedures are then as follows.

Householder Weight Procedure (HW): The LHH weight is the unbiased weight of that individual who was the householder at t_1. (As an alternative, householder could be replaced by principal person in the description of the procedure and in the subsequent discussion).

Average of Spouses Weights Procedure (ASW): For any LHH which at t_1 was not a married-couple household, the LHH weight is the same as for HW. For a married-couple LHH, the weight is the mean of the weights of the householder and spouse of the householder at t_1 with the following exception. If only one member of the couple was in the Wave 1 cross-sectional universe then the LHH weight is the weight of that spouse.

From the discussion in [3], to show that a weighting procedure provides unbiased estimates for a specific universe it suffices to show that the expected value of the weight is 1 for each LHH in the universe and 0 for all other LHH's. It can readily be shown that this statement is true for these two procedures, where the universe for the HW procedure is all LHH's for which the householder at t_1 was in the Wave 1 cross-sectional universe, while the universe for the ASW procedure is slightly larger, including also all married-couple households for which the spouse but not the householder at t_1 was in the Wave 1 cross-sectional universe.

Note that these universes include part-interval LHH's as well as LHH's in existence for an entire interval for which estimates are desired. Also note that these procedures can also

be used for unbiased estimation of more restrictive universes than those stated by simply zero weighting LHH's not in the more restrictive universes. On the other hand, estimates for universes that include LHH's not in the universes stated in the previous paragraph would generally be biased, since such LHH's are not reflected in the estimates.

In choosing between these two weighting procedures, first observe that they would yield different weights only for LHH's that began after Wave 1 and began as married-couple HH's. ASW has the advantage that it would assign a positive weight to such LHH's if either member of the couple at t_1 was an original sample person, while HW would assign a positive weight only if the householder was an original sample person. Thus, ASW would include more households in the estimation than HW and should result in estimates with lower variances. Furthermore, as previously noted, the largest universe for which unbiased estimates can be obtained is slightly larger for ASW. Since ASW has no obvious disadvantages it is the recommended procedure.

2.3 Operational Problems

For any LHH definition there are many types of weighting procedures that would yield unbiased estimates in theory. However, in practice, the problem arises that not all the information to produce such estimates is always available, even assuming, as we unrealistically do in this section, that there is perfect frame coverage and no nonresponse. This problem was mentioned in the Introduction and had been previously discussed

in [1], [2] and [3]. Specifically, in order to obtain unbiased estimates it is necessary that the definition, the weighting procedure and the operational procedures result in the following three conditions being satisfied.

1. Each LHH with a positive weight is interviewed for each month in $[t_1, t_2]$, and thus all the subject-matter data needed in the estimation is collected.

2. Sufficient information is available to determine the weight of each LHH.

3. For each LHH with a positive weight, sufficient information is available to determine t_1 and t_2.

With the original SIPP operational procedures none of these conditions are satisified for the LHH definition for either HW or ASW. However, by implementing two changes, all three conditions would be satisfied for both procedures. The first change has already been implemented and the second change is under consideration. We will first state these two changes and then explain why these three conditions would be satisfied with these changes. The necessity of these changes should become apparent by this explanation.

The first change is for the case of a married-couple HH, A_t, in sample at month t, in which one spouse, A, was in a sample household A_{t-1} at month t-1 and the other, B, was in a nonsample household B_{t-1}, at month t-1. From the last sentence in the paragraph on household continuation in Appendix A, in that situation A_t could either be the continuation of A_{t-1}, or the continuation of B_{t-1}, or A_t could be a newly formed LHH at month

t. If A_t was the continuation of A_{t-1}, this could be known without any operational changes since A_{t-1} was in sample. However, certain additional retrospective questions would be necessary to determine if A_t was the continuation of B_{t-1}. Specifically it would be necessary to know if B_{t-1} was a family HH with B as a householder at month t-1, and if so, whether the relatives of B that were in both B_{t-1} and A_t constituted a majority of the relatives of A and B that were in A_t. A set of questions more than sufficient for this purpose were written by Donald Hernandez (Population Division/Census Bureau) (Attachment B) and implemented.

The second change is that if a married-couple HH was in sample at any month t and one of the spouses was followed throughout $[t, t_F]$, then the other spouse should also be followed if they split (even if the person was not an original sample person), and an interview obtained for the individual's HH for each month in $[t, t_F]$, assuming the individual remained in the universe. This would include anyone who at month t was married to an original sample person, or married to a person who previously had been married to and living with an original sample person, etc. The purpose of this change is to insure that all LHH's with positive weights would be interviewed throughout $[t_1, t_2]$. To cite an example, consider the case where A and B married at month t, with A being the householder. Prior to month t, A and B had been living alone. A was the only original sample person. A and B had a child, C, and later separated at month t', with B and C remaining together. Then according to the LHH

definition, a LHH was formed at month t, consisting of A and B, that would have a positive weight with either the HW or ASW procedures. This LHH continued through month t′, with B and C as members but would only have been followed at month t′ if this change was implemented.

It will now be shown that with these operational changes, the three conditions previously stated for obtaining unbiased estmates would be satisfied for HW and ASW for the LHH definition.

To show 1. it will be demonstrated by induction on t that if a LHH has a positive weight, then the householder and spouse (if present) for each month $t \in [t_1, t_2]$ were followed throughout $[t, t_F]$, and hence the LHH was in sample throughout $[t_1, t_2]$. First observe that for month t_1, the householder (or the spouse in a married-couple household with the ASW procedure) was an original sample person and hence by the modified following rules the householder and spouse were followed. Next it will be shown that if the householder and spouse (if present) were followed at month t, for $t < t_2$, then this was also true at month t+1. To do this, note that by the conditions of the LHH definition, either the householder at month t+1 was the householder or spouse at month t, and consequently followed by the inductive hypothesis, or the householder at month t+1 was married to and living with at month t+1 the householder or spouse at month t, and consequently followed by the modified following rules and the inductive hypothesis. The identical argument also applies to the spouse of the householder at month t+1.

To show 2. and 3. it will first be established that t_1 can be determined for any LHH with a positive weight. Since such a LHH would have been in sample at t_1 and the weights of the householder and spouse (if present) would thus be known, knowledge of t_1 for each such household is sufficient to show 2. and is half of what is needed to show 3. To determine t_1 for each LHH with a positive weight it is sufficient to determine whether A_t was a newly formed HH for each A_t that was in the cross-sectional sample at month t and for which either the householder or spouse was an original sample person. Now the following two possibilities exists for A_t:

a. The householder of A_t was in sample in month t-1, and if A_t was a married-couple HH then the spouse was also in sample in month t-1.

b. A_t was a married-couple household that met the conditions described for the first change in operatonal procedures.

If a. holds then by the conditions of the LHH definition either:

i. A_t was a continuation of a sample household A_{t-1} at month t-1;

or

ii. A_t was a newly formed HH at month t.

If b. holds there is the additional possibility that

iii. A_t was the continuation of B_{t-1} at month t-1.

Now if i. holds this would always be known, since the composition of A_{t-1} and A_t are sufficient to determine continuity for the LHH definition. Furthermore, if iii. holds this would also be known,

but only because of the first operational change. Since ii. is the only remaining possibility, it would also always be known when ii. holds.

Finally, it will be shown that t_2 can be determined for any LHH with positive weight, which will fully establish 3. This is equivalent to being able to ascertain for any such LHH with household composition denoted by A_t at month t, whether A_t had a continuation A_{t+1} at month t+1. However, if A_{t+1} was a continuation of A_t, this can be ascertained since both A_t and A_{t+1} would have been in sample by 1.

2.4 Other Unbiased Weighting Procedures

In [3], four other unbiased weighting procedures are discussed. In this author's opinion the only one that is a realistic alternative to HW and ASW is the Beginning Date of Household Procedure (BH). This procedure assigns to each LHH the mean of the unbiased weights of the (adult) individuals who were in the LHH at t_1 and were in the Wave 1 universe. This procedure has two advantages over HW and ASW. It assigns positive weights to a larger set of LHH's, namely all households which at t_1 contained at least one original sample person. Furthermore, it enables unbiased estimates to be made for a slightly larger universe, namely all LHH's that at t_1 contained at least one person who was in the Wave 1 universe. However, it requires more changes in operational procedures. The use of BH would require retrospective questions to be asked of anyone who was a householder or spouse when they first entered sample if this

occurred after Wave 1. It would also require that anyone at month t who was either a householder or spouse of a sample HH to be followed throughout $[t, t_F]$. Both of these requirements would apply to householders of any type of HH, not only married-couple HH's.

The other three weighting procedures described in [3], Beginning Date of Interval (BI), Continuous Household Members (CM) and Average Cross-Sectional Household Weight (AW), should not be given serious consideration in this author's opinion. The primary advantage of BI over BH, which it resembles, is that it does not require retrospective questions when used for a restricted universe which does not include part-interval LHH's. However, for SIPP, where it is understood that estimates are required for part-interval LHH's, this advantage disappears. CM, as noted in [3], is not usable at all for universes which include part-interval households. Finally, AW, among other problems, requires subject-matter data for some LHH's for time periods before the LHH came into sample. Since it would not be realistic to attempt to obtain all this data retrospectively, AW should only be considered if a sufficiently accurate missing data adjustment procedure could be developed.

3. WEIGHTING ADJUSTMENTS

3.1 Preliminaries

Further notation and terminology that will be used in this section is presented here.

Each LHH with a positive unbiased weight for an interval is classified as an interviewed LHH if an interview was obtained for

each month in the interval that the LHH was in existenc otherwise it is classified as a noninterviewed LHH. (R ence will also be made in the section to interv nonint ie for specific month or months. The reader should be care note the distinction.)

A LHH is an initial LHH if $t_1 = t_B$. Otherwise i a subsequently formed LHH.

Finally, a set B of LHH's is said to be generate by a set A of either cross-sectional HH's or LHH's if there was at least one original sample person in every member of B at t_1 who previously was in a HH in A if A is a set of cross-sectional HH's, or previously was in a LHH in A for at least one month if A is a set of LHH's.

3.2 General Concepts

There are several general concepts that motivate the proposed adjustment procedures.

First, recall that the longitudinal universe consists of the cohort of all initial LHH's plus a set of subsequently formed LHH's generated by the initial LHH's. The largest universe of LHH's for which unbiased estimates could be made is dependent, as previously noted, on the weighting procedure used, but would in general exclude, for example, subsequently formed LHH's which contained no one who was in the cross-sectional universe at t_B. For a LHH universe that excludes these LHH's, it would n be appropriate to adjust the LHH weights to obtain agreemen each month with independent cross-sectional HH estimates that include

such subsequently formed LHH's. There is also the further difficulty, which is discussed in [3] and [9], that any adjustment procedure which attempted to obtain agreement with cross-sectional estimates at more than one point in time could result in such unacceptable consequences as assigning some LHH's very large or negative weights.

Instead the following general approach is proposed for adjusting the unbiased weights of the sample LHH's. First the weights for the set of sample initial LHH's would be adjusted through a procedure, described in Section 3.3, consisting of several steps which resembles in part, but with some important differences, the procedure currently used to adjust the cross-sectional SIPP weights. The final step of this procedure would be an adjustment to independent cross-sectional estimates at t_B of number of HH's by demographic characteristics.

As for the subsequently formed sample LHH's, the weighting adjustments to the set of sample initial LHH's would also result in an adjustment to the weights of subsequently formed LHH's. However, a further adjustment to the weights of such LHH's would be necessary to compensate for noninterviewed subsequently formed LHH's generated by interviewed initial LHH's. This is described in Section 3.4.

Another general LHH weighting adjustment concept, which has also been proposed for longitudinal person estimation in SIPP, is that the final adjusted weights depend on the interval for which estimates are to be made. This is motivated by the fact that there are a considerable number of sample LHH's, both initial and

subsequently formed, which were interviewed for some but not all of their period of existence. If one final weight were used for each LHH, then only LHH's that were interviewed for their entire period of existence could be used in the estimates unless data were imputed for the missing time periods for LHH's not interviewed for their complete period of existence. In contrast, the use of final weights that vary with the time interval allow the use of all LHH's that were interviewed throughout a time interval to be used in estimates for that interval, including LHH's that were not interviewed for other time periods. This should result in gains in precision.

However, to simplify this noninterview problem, not all possible noninterview patterns will be considered. Instead it will be assumed that the noninterview pattern for each LHH and person is nested, that is noninterview for one month implies noninterview for all subsequent months. Then for any actual case for which the noninterview pattern was not nested, either missing interviews would be imputed or interview data subsequent to the first noninterview month would not be used in the estimation. Among other simplifications this assumption allows LHH weights to vary only with the final month of the interval for which estimates are made. This would be accomplished by obtaining final weights for every time interval of the form $[t_B, t_E]$, where $t_E \in [t_B, t_F]$, and then using the weights for $[t_B, t_E]$ also for any interval $[t, t_E]$ with $t_B < t \leq t_E$, with the exception that all LHH's that terminate before month t would be zero weighted.

It is also assumed that at each month either a complete interview for a LHH is obtained, or no interview is obtained. In practice, of course, partial data may be obtained for a given month, such as data from some but not all of the LHH members. Then either the missing data would be imputed, or the LHH would be considered to be a noninterviewed LHH for that month.

We now proceed to detail the proposed weighting adjustment procedures with respect to an interval $[t_B, t_E]$. Section 3.3 presents the adjustments for the set of sample initial LHH's and Section 3.4 for the set of subsequently formed LHH's.

For an alternative approach to weighting adjustment in a somewhat similar context the reader is referred to [12].

3.3 Weighting Adjustments for Sample Initial LHH's

It is understood that in this subsection all LHH's referred to are initial LHH's. The following four steps of weighting adjustment are proposed for these LHH's.

1. A noninterview adjustment to compensate for noninterviews at t_B.

2. A noninterview adjustment to compensate for subsequent noninterviews among LHH's that were interviewed at t_B.

3. A first-stage ratio adjustment to reduce the contribution to the variance arising from the sampling of PSU's.

4. A second-stage ratio adjustment which adjusts the sample estimates of number of initial LHH's with specific demographic characteristics to independently derived estimates of the number of such cross-sectional households in existence at t_B.

These types of adjustment are commonly present in estimation for demographic surveys conducted by the Census Bureau, although one noninterview adjustment is generally used. The reason for proposing two such adjustments here is that this would permit a selection of variables to use in forming adjustment cells from the extensive data collected from previous interviews for LHH's that were interviewed for at least the first wave, instead of being restricted to the limited information that is available for LHH's that were not interviewed at all. Two noninterview adjustments are currently used in SIPP cross-sectional estimation for the same reason.

Each of these four steps will now be described in more detail.

3.3.1 First Noninterview Adjustment

The first noninterview adjustment is conceptually the same as for cross-sectional estimation and it appears that the same or similar weighting cells would be appropriate. Furthermore, the adjustment factors would be computed in the same manner as for SIPP cross-sectional estimation, and if the same cells were used, the factors would be exactly the same. (This assumes that the unbiased weight for each initial LHH is the same as the corresponding cross-sectional household at t_B. Each of the weighting procedures described in Section 2 satisfies this condition). The factors would be applied to all sample HH's that were interviewed at t_B, with all other initial LHH's zero weighted.

3.3.2 Second Noninterview Adjustment

The weighting factor corresponding to this adjustment would vary not only with the weighting cell that an interviewed LHH belonged to but also the ending month, t, of the LHH, in order to redistribute the weights of noninterviewed LHH's with first noninterview month t only to interviewed LHH's still in existence at month t. For each t in the interval $[t_B, t_E]$ a factor F_{tC} would be applied to all interviewed LHH's in cell C with period of existence $[t_B, t]$, while all noninterviewed LHH's would be zero weighted. (Note that for a LHH for which t is the last month for which an interview was obtained and it is not known whether the LHH continued to exist at month t, an imputation could be performed to make this determination, and hence ascertain whether the LHH was an interviewed or a noninterviewed LHH.) To compute F_{tC}, first let I_{tC} denote the weighted count in cell C (using the weights after Step 1) of interviewed LHH's with period of existence $[t_B, t]$ and let N_{tC} denote the weighted count of noninterviewed LHH's in cell C with first noninterview month t. (Note that $N_{t_BC} = 0$ because of the first noninterview adjustment). Then let

$$F_{tC} = 1 + \sum_{i=t_B}^{t} \left(\frac{N_{iC}}{\sum_{j=i}^{t_E} I_{jC}} \right) .$$

Application of this factor redistributes the weights of all noninterviewed LHH's in cell C with first noninterview month t to all interviewed LHH's in existence at month t. Furthermore, the

sum of the weights of all interviewed LHH's in cell C after this adjustment is

$$\sum_{t=t_B}^{t_E} F_{tC} I_{tC} = \sum_{t=t_B}^{t_E} \left[1 + \sum_{i=t_B}^{t} \left(\frac{N_{iC}}{\sum_{j=i}^{t_E} I_{jC}}\right)\right] I_{tC}$$

$$= \sum_{t=t_B}^{t_E} I_{tC} + \sum_{i=t_B}^{t_E} \left[\frac{N_{iC} \left(\sum_{t=i}^{t_E} I_{tC}\right)}{\sum_{j=i}^{t_E} I_{jC}}\right]$$

$$= \sum_{t=t_B}^{t_E} I_{tC} + \sum_{i=t_B}^{t_E} N_{iC},$$

which, as desired, is the sum of the weights before this adjustment of all LHH's in cell C, both interviewed and non-interviewed.

The weighting cells for this adjustment could be similar or identical to these for the SIPP cross-sectional noninterview adjustment for subsequent waves [6]. However, the cross-sectional adjustment uses only control card information, a necessary limitation because no other information obtained during previous interviews is available on cross-sectional files. However, for longitudinal files all data collected in previous interviews are available for cell formation a. should be considered.

Note that in cross-sectional SIPP estimation, the n in-terview adjustment for subsequent waves is performed afte the first-stage ratio adjustment, since the first-stage adjus.ment is an adjustment to the Wave 1 sample. However, for the proposed

SIPP LHH weighting adjustments, the noninterview adjustment in this step, like the other adjustments in this subsection, is an adjustment to the set of sample initial LHH's. It would therefore be appropriate to perform this adjustment immediately after the adjustment for Wave 1 noninterviews and before the first-stage ratio adjustment.

3.3.3 First-Stage Ratio Adjustment

Conceptually this adjustment step is smilar to the first-stage ratio adjustment for SIPP cross-sectional estimation [8] with the following possible exception. Cross-sectionally, the race variable is determined on a person, not a household, basis and consequently the adjustment is performed separately for each individual in a HH. For LHH estimation it might be more appropriate to use the race of a predetermined individual in the household at t_B, such as the householder or principal person, to form HH adjustment cells. This would enable this ratio adjustment to become a HH adjustment applied to the set of interviewed LHH's.

3.3.4 Second-Stage Ratio Adjustment

It is proposed that the estimated number of households with specific demographic characteristics at month t_B obtained from CPS estimates using the March type family weighting [5] be the controls in this final step. Before detailing this step further we will digress to make several observations.

First, there is disagreement in the statistical community over whether the March CPS type weighting system should continue to be used to provide HH and family estimates in CPS. This question and a similar question for the Consumer Expenditure Survey are currently being researched at the Census Bureau and the Bureau of Labor Statistics [4], [13]. However, until this research is completed it is appropriate to continue to use the current system.

Secondly, a key reason for controlling SIPP estimates to CPS estimates is the expected increase in the precision of the SIPP estimates by this adjustment because of the larger sample size in CPS. To obtain an alternative set of controls, SIPP weights could first be adjusted directly using the procedure used to obtain the final weights in the March CPS system. Then an optimal linear combination of the SIPP and CPS estimates could be used as controls. Such an approach, using the combined sample of both surveys, would be expected to result in estimates with even greater precision than would be obtained using CPS estimates alone as controls, since it uses the combined sample of the two surveys.

Finally, although SIPP cross-sectional weights are also controlled to CPS estimates, the proposed adjustment procedure for LHH estmation would be quite different than the current cross-sectional procedure because of the different requirements of the two procedures. The cross-sectional weights are used for both HH and person estimation. As a result, one of the requirements placed on the weighting procedure is that the weight of the

husband in each married-couple household equal the weight of the wife in order that certain estimates be identical that logically have the same value, such as the number of husbands and number of wives that are married with spouse present. This husband-wife equalization requirement results in a complicated weighting adjustment for cross-sectional SIPP, the full consequences of which have not yet been fully researched. For LHH estimation, husband-wife weight equalization should not be a consideration, since person and HH estimation could not use the same weighting system because of differences in the LHH and longitudinal person universes. For example, a person A that was not in the Wave 1 universe, would not be in the proposed longitudinal person universe [7], and hence not represented in longitudinal person estimation. However, if A subsequently joined a LHH in the LHH universe then A would be represented in LHH estimation.

Instead, the following approach is suggested for this final adjustment step for the sample initial LHH's. The weights after the first-stage ratio adjustment would be adjusted separately for each person who at month t_B was a member of an interviewed LHH. One of the variables that would determine the adjustment cells would be relationship to householder. There would be at least two categories for this variable, with householder or spouse of householder as one category. Other variables might be age, race, sex and HH type at t_B, (such as married-couple HH, other family HH, nonfamily HH). The weighting factor for each cell would be the CPS estimate of the number of individuals in that cell divided by the SIPP estimate after the first-stage ratio

adjustment. The final HH weight would then be the weight of the householder at t_B after this adjustment for all LHH's that were not married-couple HH's at t_B, and the mean of the adjusted weights of the householder and spouse for married-couple LHH's.

The SIPP LHH estimates obtained with this set of final LHH weights would agree with the CPS estimates for the total number of LHH's at t_B. Furthermore, for LHH's that were not married-couple households at t_B, the SIPP estimate of number of LHH's with householder in a particular adjustment cell would agree with the CPS estimate. Also, the SIPP LHH estimate of total number of married-couple LHH's at t_B would be in agreement with the CPS estimate. However, if the final LHH weights are used to estimate the total number of husbands or number of wives in a specific adjustment cell, the SIPP estimates would in general not agree with the CPS estimates at t_B.

One possible question concerning this proposed adjustment is the averaging of the husband's and wife's adjusted weights to obtain the final LHH weight for each married-couple LHH. Because there is evidence of generally better coverage of women than men in the demographic surveys conducted by the Census Bureau [5], it might be thought that the wife's adjusted weight alone would be a better LHH weight. However, the weights of husbands in the CPS March system have already been adjusted to compensate for this differential undercoverage, and there is consequently no obvious reason to believe either the husband's or the wife's weight is superior to the other in SIPP after adjustment to the CPS controls. For this reason and the fact that the averaging of the

weights would tend to produce a set of weights with less variability than either the set of husbands' or wives' adjusted weights, it is suggested that the averaging be done.

3.4 Weighting Adjustments for Sample Subsequently Formed LHH's

The weighting adjustment process for these set of LHH's will be broken into the following two parts.

1. The adjustments that would be appropriate if interviewed initial LHH's generated no noninterviewed subsequently formed LHH's.

2. The additional noninterview adjustments necessary because the assumption in 1. is not true.

The proposed approach to the first part is to associate a month t_B adjusted weight (a terminology that will become clear later) to each person who at t_B was a member of an interviewed initial LHH and then apply the ASW or an alternative weighting procedure with the month t_B adjusted weight used instead of the unbiased weight for each person. The month t_B adjusted weight for each person can be taken to be either the final LHH weight for the individual's initial LHH or the adjusted person weight computed in Step 4 of Section 3.3. The latter approach would appear more promising due to the differential undercoverage of individuals by demographic characteristics within an interviewed HH.

Note that this weighting adjustment for the set of sample subsequently formed LHH's would not result in the estimated number of LHH's in existence at any time other than t_B being in

agreement with independent controls. Any attempt to obtain such agreement could lead to large and negative weights as mentioned earlier.

The second part of the adjustment for subsequently formed sample LHH's presents serious complications that would not be found in SIPP longitudinal person estimation for example. To illustrate, consider the case of a sample initial LHH that moved at month t and was not followed. Prior to the move the LHH contained five people, but no information is available concerning the composition after the move. Then at one extreme each of these five people might have been living alone at month t, in which case the initial LHH generated five new LHH's at month t. At the other extreme these five people might have remained together, in which case there were no new LHH's at month t generated by the initial LHH. Furthermore, the weight of any new LHH's would in general not be known. For example, with the ASW procedure, if one of these people was living alone at month t the weight of this newly formed LHH would be that persons's month t_3 adjusted weight, together with further adjustments to be described. However, if that person instead formed a two person LHH by marrying a person who was not an original sample person, the weight would be half as much. Finally, if the person became part of a LHH in which the householder and spouse (if present) were not original sample persons then the LHH would be zero weighted. Thus, in addition to the problem of missing data - matter data, noninterviews after the first wave in the context of LHH estimation entail the additional problems of determining the

number of noninterviewed analytic units and their weights. It is envisioned that these problems would have to be handled by some form of imputation procedure.

Once this imputation is performed, it is proposed that the LHH weights for the set of sample subsequently formed LHH's be adjusted through a sequence of noninterview adjustments to compensate for noninterviewed LHH's generated by interviewed initial LHH's. For each month t after t_B a noninterview adjustment factor $f_{t,H}$ would be applied to each interviewed LHH, H, formed at month t. This factor would be computed by using recursion on t as follows. For each month $i \in (t_B, t)$ any interviewed LHH, H_i that was formed at month i would have previously received as noninterview adjustment factor f_{i,H_i}. This factor would also have been applied to each original sample person in H_i at month i. Consequently, at month t-1 each original sample person interviewed at month t-1 would have a month t-1 adjusted person weight of the form

$$W \prod_{i=t_{B+1}}^{t-1} g_{i,H_i}$$

where W is the person's month t_B adjusted weight, H_i is the person's LHH for month i and

$$g_{i,H_i} = \begin{cases} f_{t,H_i} & \text{if } H_i \text{ was formed at month i,} \\ 1 & \text{otherwise.} \end{cases}$$

Thus a noninterview adjustment factor would be applied to each original sample person for each month after t_B, that the person became a member of a newly formed interviewed LHH. Now to compute $f_{t,H}$, first compute, using the set of month t-1 adjusted

person weights, a LHH weight for each LHH formed at month t, both interviewed and noninterviewed, generated by the LHH's interviewed at month t-1. (This is where the recursion occurs.) For example, if the ASW procedure is used, this weight would be as described for that procedure with each person's month t-1 adjusted weight replacing the unbiased weight. $f_{t,H}$ is then the weighted count (using the weights just described) of all LHH's formed at month t and in the same adjustment cell as H, both interviewed and noninterviewed, generated by the set of HH's interviewed at month t-1, divided by the weighted count of interviewed LHH's formed at month t in this adjustment cell. The final LHH weight for H would then be the product of $f_{t,H}$ and the LHH weight computed using the set of month t-1 adjusted person weights. Note that if ASW, or alternately HW, is used then the final LHH weight is also the same weight as would be obtained by applying this procedure with each person's month t adjusted weight replacing the unbiased weight.

There are two principal motivations for the noninterview adjustment procedure that has just been described. First, at least ideally, the weights of noninterviewed LHH's formed at month t and generated by interviewed LHH's existing at month t-1 should be redistributed only to interviewed LHH's in the same adjustment cell that were also formed at month t. Secondly, there exist noninterviewed LHH's formed at month t not generated by the set of HH's interviewed at month t-1. This set of noninterviewed LHH's is not compensated for by the month t factors, but is compensated for by the factors for the prior months, which

are part of the final LHH weights for the set of interviewed LHH's formed at month t because of the recursion.

In practice there would be at least one major difficulty in computing the $f_{t,H}$ factor using the method just outlined. In general, the number of interviewed LHH's formed at month t may be too small to form adjustment cells containing a sufficient number of cases. Consequently, some compromise would undoubtedly be necessary to the principal that noninterviewed LHH's formed at month t and generated by interviewed LHH's existing at month t-1 should have their weights distributed only to interviewed LHH's formed at month t.

REFERENCES

[1] Ernst, Lawrence R., "Preliminary Thoughts on Un[illegible]ased Longitudinal Household Estimation, U.S. Bureau [illegible] the Census report, 1983.

[2] _____, "The Impact of Conditions in a Longitu[illegible]inal Household Definition on the Ability to Produce [illegible] Longitudinal Estimates in SIPP," U.S. Bureau o[illegible]he [illegible]nsus report, 1985.

[3] _____, Hubble, David L., and Judkins, Davi[illegible], "Longitudinal Family and Household Estimatio[illegible] SIPP," American Statistical Association - Proceedings of the Section on Survey Research Methods, 1984, 682-687.

[4] Gbur, Edward, "A Method of Obtaining Group Weights from Individual Weights in a Survey," American Statistical Association - Proceedings of the Section on Survey Research, 1985, to appear.

[5] Hanson, Robert H., The Current Population Survey - Design and Methodology, U.S. Bureau of the Census, Technical Paper 40, 1978.

[6] Jones, Charles D., "SIPP Weighting - Subsequent Wave Cross-Sectional - Revised," U.S. Bureau of the Census memorandum to Thomas C. Walsh, October 16, 1984.

[7] _____, "Overview of SIPP Longitudinal Weighting - Revision," U.S. Bureau of the Census memorandum to Roger A. Herriot, January 24, 1986.

[8] _____, "SIPP 85: Cross-Sectional Weighting Specifications for Wave 1," U.S. Bureau of the Census memorandum to Thomas C. Walsh, July 2, 1985.

[9] Judkins, D. R., Hubble, D. L. Dorsch, J. A., McMillen, D. B, and Ernst, L. R., "Weighting of Persons for SIPP Longitudinal Tabulations," American Statistical Association - Proceedings of the Section on Survey Research Methods, 1984, 676-681.

[10] Nelson, Dawn D., McMillen, David B., and Kasr[illegible]yk, Daniel, "An Overview of the Survey of Income and Program [illegible]ticipation: Update 1," SIPP Working Paper Series, No. 401 [illegible].S. Bureau of the Census, 1985.

[11] Singh, Rajendra P., "Some Thoughts on Longitudi[illegible] Household Weighting for SIPP," U.S. Bureau of the Census [illegible]orandum to Gary M. Shapiro, April 3, 1984.

[12] Whitmore, Roy W., Cox, Brenda G., and Folsom, Ra[illegible]h E., "Family Unit Weighting Methodology for the Natur[illegible] Household Survey Component of the National Medical Care Utilization and Expenditure Survey," Research Triangle Institute report, 1982.

[13] Zieschang, Kimberly D., "Sample Weighting Methods and Estimation of Totals in the Consumer Expenditure Survey," Bureau of Labor Statistics report, 1985.

Attachment A

DRAFT: July 2, 1985
Donald J. Hernandez

THE CONTINUATION, DISSOLUTION, AND FORMATION OF HOUSEHOLDS LONGITUDINAL DEFINITION FOR SIPP

A household continues, dissolves, or forms depending upon the nature of changes from one month to the next in the living arrangements of householders and their spouses, as defined below.

Household Continuation

A household continues from one month to the next, if one of three conditions is met for at least one person who is either the householder or the householder's spouse during both months. First, during both months the person maintained a household with no other persons present. Second, during both months the person maintained a household with one or more additional persons present, none of whom were related to the householder. Third, during both months the person lived with at least one other relative who was present during both months. If the latter condition is met by two persons who were in different households during one of the months, then select one of these persons as the one in terms of whom the continuing household is defined by applying the following rule: select the person living with a specific set of own family members in the household during both months who constitute a majority of the householder's or spouse's family members in both months.*

Household Dissolution

A household dissolves between one month and the next month, if the household existed during the first month but it did not continue from the first month to the second month.

Household Formation

A household is formed during a month, if the household existed during the month but it did not continue from the preceding month.

Further Considerations

A household is classified as continuing for a period of more than two months by cumulating month to month decisions. For example, a household that continues from month 1 to month 2 and from month 2 to month 3 is defined as continuing through the entire period. In addition, a nonfamily household continues as a family household between month 1 and month 2 if the change in household type occurs because two unmarried persons in the household in month 1 become married to each other and continue to share the same household in month 2.

*One assumption necessary to make this definition complete is that no individual can be married to and living with one person at month t and another person at month t+1.

Attachment B

OMB No. 0607-0425; Approval Expires August 31, 1987

FORM **SIPP-5011**
6-11-85

U.S. DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS

LONGITUDINAL HOUSEHOLD CONTROL FORM

SURVEY OF INCOME AND PROGRAM PARTICIPATION

1985 PANEL

NOTICE — Your report to the Census Bureau is **confidential** by law (title 13, U.S. Code). It may be seen only by sworn Census employees and may be used only for statistical purposes.

a. R.O. code (cc item 1)	**b.** Control number (cc item 2) PSU	Segment	Serial	Sample	Check digit	**c.** Add. ID (cc item 3)	**d.** Wave this form filled

e. Entry Address ID/Person Number (cc items 17–18)

f. Person name — *Last, first, middle, maiden* (cc item 19a)

INTERVIEWER INSTRUCTION ➧ *WHENEVER YOU MARK THE "NO" BOX IN CONTROL CARD ITEM 21F, FILL O[illegible] OF THESE FORMS FOR EACH PERSON WHO HAS A PERSON NUMBER EQUAL TO 200+*

CHECK ITEM A During the previous wave, was . . . the reference person or the spouse of the reference person (cc item 19b equals 1, 2, or 3)?

0001
1 ☐ Yes
2 ☐ No — *SKIP to next person*

CHECK ITEM B Is . . .'s left code in cc item 23 in the range 6–12, 26–31?

0002
1 ☐ Yes — *SKIP to item 1*
2 ☐ No

CHECK ITEM C Does the current wave entry in cc item 36b equal 16 or 23–26?

0003
1 ☐ Yes
2 ☐ No — *SKIP to next person*

ASK OR VERIFY —

1. Is . . . now living in any of the kinds of places listed on this card?
(SHOW FLASHCARD V)

0004
1 ☐ Yes — *SKIP to Check Item D*
2 ☐ No
x1 ☐ DK
x2 ☐ Ref.

2. Is . . . now living alone?

0005
1 ☐ Yes — *SKIP to Check Item D*
2 ☐ No
x1 ☐ DK
x2 ☐ Ref.

3. Is . . . the owner or the renter of . . .'s current residence?

0006
1 ☐ Yes — *SKIP to item 5*
2 ☐ No
x1 ☐ DK — *SKIP to item 5*
x2 ☐ Ref. — *SKIP to item 5*

4. How is . . . related to the person who owns or rents . . .'s current residence?

0007
1 ☐ Spouse
2 ☐ Other — *SKIP to Check Item D*
x1 ☐ DK
x2 ☐ Ref.

5. How many of . . .'s relatives are living with . . . now?

0008 ☐☐ Enter number of persons
x3 ☐ None
x1 ☐ DK
x2 ☐ Ref.
} *SKIP to Check Item D*

6. How many of these relatives lived with . . . when . . . was living with *(First 100 level person listed in cc item 19)*?

0009 ☐☐ Enter numb[illegible]f perso[illegible]
x3 ☐ None
x1 ☐ DK
x2 ☐ Ref.

CHECK ITEM D Is the respondent to this form a current household member?

0010
1 ☐ Yes — *Enter person number* — 0011 ☐☐☐
2 ☐ No

CHECK ITEM E Enter the current wave household interview status from cc item 36b.

0012 ☐☐