

BUREAU OF THE CENSUS  
STATISTICAL RESEARCH DIVISION REPORT SERIES  
SRD Research Report Number: Census/SRD/RR-85/16

THE IMPACT OF CONDITIONS IN A LONGITUDINAL  
HOUSEHOLD DEFINITION ON THE ABILITY TO PRODUCE  
UNBIASED LONGITUDINAL ESTIMATES IN SIPP

by

Lawrence R. Ernst  
Statistical Research Division  
Bureau of the Census

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Recommended by: Paul Biemer

Report completed: February, 1985

Report issued: February, 1985

## 1. INTRODUCTION

Three key steps in the development of procedures for obtaining unbiased longitudinal household estimates of population totals in the Survey of Income and Program Participation are:

1. Select a definition of household continuity.
2. Develop a method for creating longitudinal household files with the definition selected in Step 1.
3. Determine a procedure for obtaining a set of weights for the longitudinal household files created in Step 2 which yield unbiased estimates.

(In this paper adjustments to the unbiased estimates are not discussed. Also, such problems as imperfections in the sampling frame, nonresponse, and response errors, which in general result in truly unbiased estimates being unobtainable in practice, are ignored. It is assumed that the longitudinal universe is a partition of the union of all households in the cross-sectional universe for one or more months during the interval under study, except those cross-sectional households which contain no people who were in the cross-sectional universe at the time of the first interview.)

Ideally, Step 1 above should be undertaken based only on subject-matter concepts. Unfortunately, with the current SIPP operational procedures the ability to perform Steps 2 and 3 is very much dependent on the longitudinal household definition adopted, and therefore it is necessary to be aware of the potential impact on Steps 2 and 3 of any condition that is part of a longitudinal household definition under consideration.

In Section 2 of this paper, five potential problems that can interfere with Steps 2 and 3 are presented. In Section 3 conditions are given that are sufficient for avoiding each of these problems. The paper concludes with a

discussion in Section 4 of the alternatives in terms of changing operational procedures or accepting biased estimates if a definition is adopted which results in the problems described in Section 2.

## 2. PROBLEMS ASSOCIATED WITH CREATING LONGITUDINAL HOUSEHOLD FILES AND OBTAINING UNBIASED WEIGHTS

The following are three major potential problems associated with the creation of longitudinal household files.

A. Longitudinal households for which data is incomplete for a time interval of interest because the household existed before entering sample or after leaving sample.

B. Longitudinal households for which it cannot be determined whether the household existed before entering sample or after leaving sample.

C. Inability to determine in some cases if a household  $A_{t+1}$  at month  $t+1$  will be the continuation of a household  $A_t$  at month  $t$  even when both  $A_t$  and  $A_{t+1}$  are in sample. (In this paper it is assumed, for simplicity of presentation, that a household is considered continuous over an interval of  $n$  months if it is continuous for each of the  $n-1$  corresponding pairs of consecutive months in the interval. However, the discussion also applies to some other approaches to continuity over an interval.)

In addition to Problems A, B, and C there is a potential fourth problem which can interfere with the ability to obtain a set of weights which yield unbiased estimates, namely:

D. The weights associated with some longitudinal households depend on information about the household before it entered sample or after it left sample, and which is consequently not available.

Fortunately, there is a potential fifth problem which if it is avoided makes it unnecessary to be concerned with problems A and B in producing unbiased longitudinal estimates, namely:

E. Positive weights must be assigned to some households with problems A or B.

Thus, if problem E is avoided then any longitudinal household with problems A or B can be assigned a weight of zero and hence need not be used in the estimation process. Therefore, any longitudinal household definition for which there exists an unbiased weighting procedure which avoids problems C, D and E allows unbiased estimates to be obtained with the current operational procedures.

### 3. CONDITIONS FOR AVOIDING PROBLEMS IN PRODUCING UNBIASED LONGITUDINAL ESTIMATES

For each of the problems described in the previous section, conditions in a longitudinal household definition which are sufficient to avoid the particular problem are presented.

Problems A and B. It appears that the only longitudinal household definition that avoids these two problems is the no change definition, according to which a household is continuous over a time interval if there is no change in household composition during the time interval (or alternatively no change among original sample people, that is those individuals who are to be followed throughout the life of the panel.) It is clear that this definition avoids these two problems. For if a household enters sample after the first wave, then an original sample person must join the sample at that time and hence the household would be newly formed at that time under the no change definition. Furthermore, then during its entire period of existence the household will contain original sample people and data will therefore be collected. Finally, under this definition if a change in composition takes place, this will be known and the household will be dissolved at that time. Thus, neither problems A or B can occur with the no change definition.

On the other hand, with any other definition consider a situation where an original sample person takes up residence with a set of individuals who are not original sample persons. (Such individuals are referred to as associated sample persons in this paper.) In general, it cannot always be determined if this is a continuing or a newly formed household without obtaining retrospective information, not presently being asked, from the associated sample persons. Furthermore, if the household existed before the original sample person joins, then the data for the household would be incomplete. Thus, problems A and B can both arise. Similarly, if the original sample person subsequently leaves the household, then since the additional sample people will not continue to be interviewed, it will not always be possible to determine if the household will continue to exist, and if it does continue no further data will be collected.

Problem C. There is a relatively broad class of definition that avoid Problem C. In general, Problem C is avoided if and only if for each pair of households  $A_t$  at month  $t$  and  $A_{t+1}$  at month  $t+1$ , the determination of whether  $A_{t+1}$  will be the (unique) continuation of  $A_t$  and  $A_t$  is the (unique) antecedent of  $A_{t+1}$  depends only on the set of people in  $A_t$  and  $A_{t+1}$ . Thus, if the determination of whether  $A_{t+1}$  will be the continuation of  $A_t$  depends on the composition of some other household  $B_t$  at month  $t$  (or  $B_{t+1}$  at month  $t+1$ ), problem C may arise. This is because  $B_t$  (or  $B_{t+1}$ ) may contain no original sample people, in which case its composition is not known.

As an illustration of a condition in a longitudinal household definition which would lead to Problem C, consider a plurality rule, that is the requirement that  $A_{t+1}$  must be a household with the largest number of people from  $A_t$  in order for  $A_{t+1}$  to be a candidate for a continuation of  $A_t$ . For example, if  $A_t$  contains five people and exactly two of these people are in  $A_{t+1}$ , while the other three are not residing with any original sample people at month  $t+1$ , then

it cannot be determined if the plurality rule will be satisfied, since it will not be known how many of the other three people in  $A_t$  will be members of the same household at month  $t+1$ .

As an illustration of a condition for which Problem C would not arise, consider the reciprocal majority rule, that is the requirement that in order for  $A_{t+1}$  to be the candidate for the continuation of  $A_t$  a majority (strictly greater than one-half) of the members of  $A_t$  will be in  $A_{t+1}$  and a majority of the members of  $A_{t+1}$  are in  $A_t$ . For example, if  $A_t$  contains five members and  $A_{t+1}$  contains seven members then this condition will be satisfied if and only if at least four members of  $A_t$  are in  $A_{t+1}$ . Thus, the condition can be verified with only knowledge of the composition of  $A_t$  and  $A_{t+1}$ .

Some other examples of conditions in a longitudinal household definition which do not lead to Problem C are:

1.  $A_t$  and  $A_{t+1}$  have identical household members.
2.  $A_t$  and  $A_{t+1}$  have the same householder (or, alternatively, principal person).
3.  $A_t$  and  $A_{t+1}$  have at least two (or, alternatively, any fixed number persons in common).
4.  $A_t$  and  $A_{t+1}$  have at least two (or any fixed number) related persons in common.
5.  $A_t$  and  $A_{t+1}$  are either both family households or both nonfamily households.

Note, however, that for Examples 3, 4, and 5, if the condition is satisfied this does not, in general, guarantee that  $A_{t+1}$  will be the unique candidate for the continuation of  $A_t$  and that  $A_t$  is the unique candidate for the antecedent of  $A_{t+1}$ . Further conditions are necessary in the definition to ensure uniqueness.

As an additional example of an aspect of a longitudinal household definition

that would lead to Problem C, consider the requirement that if more than one household qualifies as the continuation of  $A_t$  (or the antecedent of  $A_{t+1}$ ) according to the other conditions in the definition, then the choice among all that qualify will be made by selecting among them randomly with equal probability. The reason that this requirement results in Problem C is that not all households that qualify as the continuation (or antecedent) will necessarily contain original sample people, and consequently not all the households from which the selection should be made will necessarily be known.

Problems D and E. Both of these problems can be avoided if the definition includes a condition which can be expressed in the following form. Associated with each household  $A_t$  in existence at time  $t$  is a subset  $S_{A_t}$  of  $A_t$  such that  $S_{A_{t+1}} = S_{A_t}$  is a necessary condition for  $A_{t+1}$  to be the continuation of  $A_t$ . (In this case the weight that will be assigned is the average of the first wave weights of all members of this set that were in the universe during the first wave. The first wave weight is the reciprocal of the probability of selection for all original sample people and zero for all associated sample people. See Ernst, Hubble, and Judkins (1984) for further details.) This rather abstract concept will be illustrated by several examples.

First, consider the condition that  $A_t$  and  $A_{t+1}$  must have the same householder (or, alternatively, principal person) for  $A_{t+1}$  to be the candidate for continuation of  $A_t$ . In this example  $S_{A_t}$  is a set containing exactly one person, namely the householder (principal person) of  $A_t$ , and the requirement that  $S_{A_{t+1}} = S_{A_t}$  amounts to simply restating the same householder (principal person) condition.

A second example is the requirement that if  $A_t$  is a married-couple household then  $A_{t+1}$  must have the same householder and spouse as  $A_t$  for it to be the candidate for the continuation of  $A_t$ , while for all other types of households  $A_{t+1}$  must have the same householder as  $A_t$ . In this case  $S_{A_t}$  consists of the

householder and spouse of  $A_t$  in a married-couple household, and the householder of  $A_t$  for all other households.

A third example is the no change definition, for which  $S_{A_t} = A_t$ .

A less stringent condition is sufficient for avoiding Problems D and E when the universe of interest is restricted to the set of longitudinal households that are in existence throughout the entire interval for which estimates are desired. In this case, provided that the definition requires that at least one person be present throughout the entire interval of interest, problems D and E can both be avoided. (The weighting procedure that accomplishes this is to take the average of the first wave weights of all individuals who are members of the household for the entire interval and were in the universe during the first wave. Note though, that when using this weighting procedure the household weight may vary with the time interval under consideration. See Ernst, Hubble, and Judkins (1984) for further details.)

#### Avoiding All Problems in Obtaining Unbiased Estimates

In summary, to avoid having any of these five problems affect the ability to produce unbiased estimates, first include the type of condition in the longitudinal household definition which avoids Problems D and E. Then even if there are households with Problems A and B they need not be used in the estimation. Additional conditions can also be included in the definition, and if these additional conditions do not introduce Problem C, then there would be no estimation difficulties.

For example, the following is a longitudinal household definition for which unbiased estimates can be obtained.  $A_{t+1}$  at time  $t+1$  will be the continuation of  $A_t$  at time  $t$  if  $A_t$  and  $A_{t+1}$  have the same householder and the reciprocal majority rule holds. In this definition the same householder



requirement guarantees that Problems D and E would not arise, and that Problems A and B need not be of concern in estimation. Furthermore, the additional condition that the reciprocal majority rule holds does not introduce Problem C.

Inclusion of a Condition on the Presence of Original Sample  
People in a Longitudinal Household Definition

A condition that has appeared in some proposed longitudinal household definitions in an attempt to avoid the previously mentioned problems is the requirement that an original sample person must be present throughout a time interval in order for the household to be a candidate for continuity, or some variation of this condition. It is this author's belief that such a condition cannot actually be part of a longitudinal household definition, which is a population, not a sample related concept. That is, continuity of a household should be independent of the outcome of any sampling process. In reality then, this type of condition would be part of the estimation process rather than part of the definition, since what it would accomplish is place restrictions on the sample units to be used in the estimation. However, in general it does not appear possible to produce unbiased estimates with the restrictions imposed. For example, such restrictions can result in some longitudinal households having no chance of being used in the estimation process, which immediately guarantees that unbiased estimates cannot be obtained.

However, note that although a condition on the presence of original sample people should not be part of a longitudinal household definition, for some definitions it is possible to develop unbiased weighting procedures which only assign positive weights to longitudinal household which have original sample people present throughout the time interval for which estimates are being produced. The definitions for which this can be done are precisely those

which avoid Problem E. Adding a condition on the presence of original sample people to definitions for which Problem E is present in an attempt to work around this problem will not work.

#### 4. ALTERNATIVES IF PROBLEMS IN PRODUCING UNBIASED LONGITUDINAL ESTIMATES ARE NOT AVOIDED

If a longitudinal household definition is adopted which fails to avoid all of Problems A-E with the current operational procedures, then, in general, either the operational procedures would have to be changed or a bias would be introduced in the estimates. The alternatives are discussed in this section.

If a definition results in Problem E with the current operational procedures, then in order to obtain unbiased estimates it would be necessary to change the operational procedures so information not now collected can be obtained for households for time periods before any original sample people joined and after all original sample people left. With some estimation procedures it would only be necessary to collect sufficient information to determine if a household was newly formed or continuing at the time the first original sample person entered, and to determine if it dissolved or continued when the last original sample person left. With this additional information Problems D and E would be avoided and Problems A and B need not be a factor in the estimation. However, obtaining this additional information may not always be practical. In the case when the first original sample person entered a household, retrospective questions that could become unreasonably complicated for many definitions would be required to obtain this information. In the case when the last original sample person left at month  $t+1$ , it would be necessary to follow all associated sample persons to determine the composition of their households at month  $t+1$ , and it may also be necessary to obtain retrospective

information from any individuals who at month  $t+1$  became members of a household which contained at month  $t+1$  people who at month  $t$  were members of the household whose continuity is to be determined.

If a definition additionally results in Problem C with the current operational procedures, then in order to obtain unbiased estimates it would also sometimes be necessary to follow associated sample people even when original sample people remain in the household, for the purpose of verifying that the household continues to exist.

If a longitudinal household definition and set of operational procedures are adopted which together make it impossible to obtain unbiased estimates, then a key question is how large are the resulting biases. Certainly no definitive answers can be provided until SIPP has been operating for a long enough period for a meaningful empirical study of this question to be made. However, it would be expected that for many statistics the bias would not be vary large because of the stability of most households over the length of a SIPP panel. It also might be possible to alleviate the bias problem by some type of missing data procedure for the households that cause difficulty in the estimation process. However, for certain types of analyses it is precisely these problem households, that is the ones with compositional changes, that are of primary interest, and such analyses may be severely hampered by the inability to obtain unbiased estimates.

## REFERENCE

Ernst, Lawrence R., Hubble, David L., and Judkins, David R., (1984), "Longitudinal Family and Household Estimation in SIPP," American Statistical Association - Proceedings of the Section on Survey Research Methods, to appear.