# OBTAINING CROSS-SECTIONAL ESTIMATES FROM A LONGITUDINAL SURVEY:

# EXPERIENCES OF THE INCOME SURVEY DEVELOPMENT PROGRAM

Hertz Huang, Bureau of the Census

## I. INTRODUCTION

In 1975 the Secretary of the Department of Health, Education and Welfare (The Department of Health and Human Services (HHS) predecessor agency) authorized a program, the Income Survey Development Program (ISDP), to resolve technical and operational issues for a major new survey --the Survey of Income and Program Participation (SIPP). Much of the work of the ISDP centered around four experimental field tests that were conducted in collaboration with the Bureau of the Census to examine different concepts, procedures, questionnaires, recall periods, etc. Two of the tests were restricted to a small number of geographic sites; the other two were nationwide. Of the two nationwide tests, the more important data collection was called the 1979 Research Panel. This panel consisted of nationally representative samples which provided a vehicle for feasibility tests and controlled experiments of alternative design features. Information concerning the ISDP may be found in Ycas and Lininger (1981), David (1983), and the survey documentation now available through the National Technical Information Service (1983).

The 1979 Research Panel was a multiple frame sample consisting of a general population (area) sample of 9300 initially designated addresses drawn from the 1976 Survey of Income and Education (SIE) and some Census Bureau's current survey reserve measures and new construction update, and two list frame samples of (a) eligible applicants for the Basic Educational Opportunity Grant (BEOG) Program and (b) blind and disabled Supplemental Security Income (SSI) recipients.

The 1979 Research Panel was a longitudinal survey consisting of six waves of interviewing. The sample was divided into three interviewing panels. The first panel was first interviewed in February 1979; the second panel was first interviewed in March; and the third panel was first interviewed in April. Each sample unit was subsequently interviewed every three months. A sample of addresses was chosen and persons living in the sample units (addresses) during the first wave of interviews were defined as original sample persons. For interviews subsequent to the first, the sample of addresses became a sample of persons; accordingly, original sample people were followed to their new addresses in subsequent interviews (with reasonable geographic constraints -- within 50 miles of any ISDP Primary Sampling Unit). Personal interviews were conducted in Wave 1 with all adults (persons sixteen years old and over) at the sampled address. These become the original sample persons. During Waves 2-6 all persons currently residing with an original sample person were interviewed. This means, for example, that if an original sample person moved to a new address with four other adults, then questionnaires were administered to everyone at the original sample person's new address. If any original sample person remained at the first wave address, anyone who moved into that address with the original sample person was also interviewed. Thus, interviews were conducted with all adults at an address as long as at least one of the adults present was an original sample person. Because of the ISDP rules, persons can be lost from sample because they move beyond the survey's boundaries; in addition, people were added to the sample because they became part of the housing unit in which the original sample person resides.

Obviously, the universe changes continuously through the life of the survey. A great deal of interest exists, however, in developing cross-sectional estimates at the time of each interview wave. In the absence of drawing a new sample at each interview, any cross-sectional estimates developed for Waves 2-6 are subject to a population coverage bias. This paper will focus only on the covered population and present some unbiased base weights for cross-sectional estimators for the non-institutionalized U.S. population represented by the longitudinal sample (the population coverage bias will remain, however, since units containing no persons who were in the universe at the time of Wave 1 cannot come into sample). Since the methodology for treating both area sample and list frame samples was needed for ISDP 1979 Research Panel, both will be described below. The estimation methods described here are directly applicable to the Survey of Income and Program Participation (SIPP), an overall description of which is found in Nelson, McMillen, and Kasprzyk (1984) and Herriot and Kasprzyk (1984).

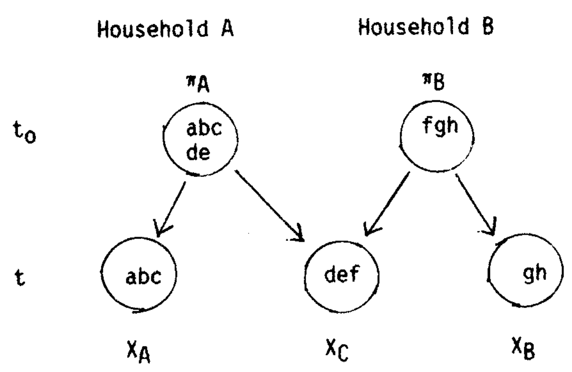
## II. THE POPULATION FOR CROSS-SECTIONAL ESTIMATES

We begin by defining the general population for which estimates are required. All households existing during the first wave of interviews (February through April 1979) are considered the initial population. Based on the rules adopted for the following individuals who move, we have essentially a longitudinal sample of persons as well as households for the initial population. Since no new sample was drawn at any subsequent interview, the sample does not completely represent the non-institutionalized U.S. population after first quarter of interview. There were persons in the following categories at the initial interview time but became part of the non-institutional population at a subsequent wave of interviewing: 1) U.S. citizens living abroad, 2) citizens of other countries who subsequently move to the U.S., 3) persons in institutions or armed forces barracks. These persons will be called the group R subpopulation which did not have chance to be selected as original sample persons. At a subsequent wave of interviews, the longitudinal sample did not include any household in which all current members were in the group R subpopulation. However, persons in the group R subpopulation who later joined households that included original persons eligible for sampling in the first wave were added to the cross-sectional universe. these persons along with newborns will be called "additions" in subsequent waves. In general, "additions" are defined as persons moving into eligible households after the first wave who were not eligible for sampling in the first wave.

## III. GENERAL CONCEPT OF CROSS-SECTIONAL ESTIMATION

Due to the procedures adopted for following movers in the 1979 Research Panel, at subsequent interviews a household could consist of members from more than one household in the universe at the time of the first wave. The inclusion probability of such a household would depend on the inclusion probabilities of the households which the members of the current household were part of at the time of the first interview. The inverse of the inclusion probability is usually used as the weight of a sample household in estimation. However, because of the sample design of the 1979 Research Panel, the inclusion probability of a household is a function of its primary sampling unit, type of sampling frames and the 1975 income of the household which occupied the housing unit during the SIE interview. Only the inclusion probability of an original sample household was feasible to calculate. The inclusion probability of an original nonsample household is almost impossible to evaluate, but such households can come into sample on later waves. Therefore, some alternative weighting procedures needed to be explored.

The idea to be presented in this discussion is very simple. We will associate observations at any given point in time with the known inclusion probabilities of the original sample households. We will split up observations belonging to a household when current household members come from more than one original household. A portion of the observation is then associated with each original household. The following example will illustrate the idea: Assume that A & B are two original households with inclusion probabilities $\pi_A$ and $\pi_B$ respectively. At the first wave of interviews, household A consists of five members, a,b,c,d, and e, and the household B consists of three members, f,g, and h. During the second wave of interviews we find that d,e, and f are living together and form a new household, called household C, while a,b, and c are still in household A and g and h are still in household B.




Two alternatives are proposed, both involving the division of household C into two parts; one part is associated with household A and the other with household B.

a) Multiplicity Approach:

Based on the number of ways (called multiplicity) that the new household C can be included in the sample, the observation (additive, such as counts, income or values) of household C (called $X_C$) is divided by the number of original households involved (two in this case) and each portion is added to the corresponding observation of household A (called $X_A$) and household B (called $X_B$). Therefore, if both households A and B are original sample households, the cross-sectional estimate, $\hat{X}$, for the total at the second wave based on these three households can be expressed as:

$$\hat{X} = \frac{1}{\pi_A}\left(X_A + \frac{1}{2}X_C\right) + \frac{1}{\pi_B}\left(X_B + \frac{1}{2}X_C\right)$$

$$= \frac{1}{\pi_A}X_A + \frac{1}{\pi_B}X_B + \left(\frac{1}{2\pi_A} + \frac{1}{2\pi_B}\right)X_C.$$

Hence, the weight for the new household c is $\frac{1}{2\pi_A} + \frac{1}{2\pi_B}$. If only household A is an original sample household, then the weight for the new household is $\frac{1}{2\pi_A}$; if only household B is an original sample household then the weight for the new household is $\frac{1}{2\pi_B}$.

b) Fair Share Approach

This approach assumes that all household members contribute equally to their household. Thus, the observation of household C is divided into appropriate portions based on the proportion of members of household C which come from each original household (2/3 from household A and 1/3 from household B in this example). Therefore, if both households A and B are original sample households, the cross-sectional estimate for the total at the second wave based on these three households is expressed as

$$\hat{X} = \frac{1}{\pi_A}\left(X_A + \frac{2}{3}X_C\right) + \frac{1}{\pi_B}\left(X_B + \frac{1}{3}X_C\right)$$

$$= \frac{1}{\pi_A}X_A + \frac{1}{\pi_B}X_B + \left(\frac{2}{3\pi_A} + \frac{1}{3\pi_B}\right)X_C.$$

Hence, the weight for the new household c is $\frac{2}{3\pi_A} + \frac{1}{3\pi_B}$. If only household A is an original sample household, then the weight for the new household is $\frac{2}{3\pi_A}$; if only household B is an original sample household then the weight for the new household is $\frac{1}{3\pi_B}$.

Since our sample was longitudinal in nature, if the universe remained constant through time, original sample persons would be a representative sample of the universe at any given point in time. Hence, using the inclusion probabilities of the original sample persons, the above estimators are unbiased (proof is given in next section). However, our feasible target population (excluding the group R subpopulation) changes through time by including the 'additions' (defined in Section II). To compensate for this, we will include these "additions" in the proposed estimators below.

## IV. PROPOSED ESTIMATORS FOR GENERAL POPULATION (AREA) FRAME

Before the estimators are given, some notation should be defined. Note that some of the defined quantities may not be observed. For the first wave of interview (time $t_0$), let

$X(t_o) = \sum_{k=1}^{N(t_o)} X_k(t_o)$ the parameter to be estimated, where $X_k(t_0)$ is the value of the characteristic for the $k^{th}$ unit (which may be a person or a household) and $N(t_0)$ is the number of units in the universe at time $t_0$;

$\alpha_k$ = 1 if unit k was in the sample at time $t_0$, k = 1, ..., $N(t_0)$
= 0 otherwise

$\pi_k$ = the probability that unit k was selected for the sample at the first wave of interview (time $t_0$)
= $P_r(\alpha_k = 1) = E(\alpha_k)$, k=1,..., $N(t_0)$ At a subsequent wave (time t), define for each household i:

$S_i$ = the total number of current residents of household i at time t

$r_i$ = the number of original eligible households from which the current residents come (does not include households from which "additions" come)

and

$S_{i1}, S_{i2}, \ldots, S_{ir_i}$ be the number of current residents in household i from each of the $r_i$ original households and $S_{ia}$ be the number of current residents from the "additions" as defined in Section II. Write

$$S_i = \sum_{j=1}^{r_i} S_{ij} + S_{ia} \quad \text{and} \quad S_{io} = \sum_{j=1}^{r_i} S_{ij}$$

Also define N(t) as the total number of units in the target population at time t, such as household units (include all the original households plus newly formed households) or other units based on a group of persons such as families or sub-families (include all sample persons interviewed nonsample persons plus "additions"). And let $X(t) = \sum_{i=1}^{N(t)} X_i(t)$ be the parameter (total) to be estimated for the target population at time t. To simplify the notation, we will replace N(t), X(t) and $X_i(t)$ by N, X and $X_i$ respectively.

Based on the general concept described in Section III, two cross-sectional estimators are proposed for the area frame to estimate the total of a characteristic of the target population at time t.

Estimator I (Multiplicity Estimator):

This estimator is based on the multiplicity of each current household

$$\hat{X}_M = \sum_{i=1}^{N} W_{Mi} X_i$$

where

$$W_{Mi} = \sum_{j=1}^{r_i} \frac{\alpha_j}{r_i \pi_j}.$$

Note that $\alpha_j$ and $\pi_j$ are associated with original households but are reindexed within each current household i. It is easily seen that

$$E(\hat{X}_M) = E\left(\sum_{i=1}^{N} W_{Mi} X_i\right) = E\left(\sum_{i=1}^{N} \sum_{j=1}^{r_i} \frac{\alpha_j}{r_i \pi_j} X_i\right)$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{r_i} \frac{E(\alpha_j)}{r_i \pi_j} X_i = \sum_{i=1}^{N} \frac{1}{r_i}\left(\sum_{j=1}^{r_i} \frac{\pi_j}{\pi_j}\right) X_i = \sum_{i=1}^{N} X_i = X$$

Therefore $\hat{X}_M$ is an unbiased estimator of X. Note also that if $\alpha_j = 0$ it is not necessary to know $\Pi_j$, so that $W_{Mi}$ can be calculated based on the selection probability only for sample units.

Estimator II (Fair Share Estimator):

This estimator is motivated by the assumption that all current household members contribute equally to the household in which they reside for the major household characteristic values, such as earnings and welfare benefits.

$$\hat{X}_F = \sum_{i=1}^{N} W_{Fi} X_i$$

where

$$W_{Fi} = \sum_{j=1}^{r_i} \frac{S_{ij}}{S_{io}} \frac{\alpha_j}{\pi_j}.$$

As in the multiplicity estimator, $\alpha_j$ and $\pi_j$ are associated with original households but are reindexed within each current household i. One can see that $\hat{X}_F$ is also an unbiased estimator of X as follows:

$$E(\hat{X}_F) = E\left(\sum_{i=1}^{N} W_{Fi} X_i\right)$$

$$= \sum_{i=1}^{N} \frac{1}{S_{io}}\left(\sum_{j=1}^{r_i} \frac{S_{ij} E(\alpha_j)}{\pi_j}\right) X_i$$

$$= \sum_{i=1}^{N} X_i = X$$

Note that if household j was not in sample at time $t_0$, it is unnecessary to know the number of current residents from original household j, $S_{ij}$, in $\hat{X}_F$ since $\alpha_j = 0$. Also note that because "additions" are not included in the weight calculations, they must be identified and excluded from using either estimator.

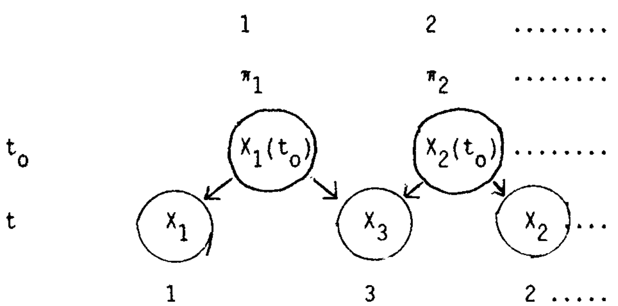
Comparison of Estimator I and Estimator II

Both Estimator I, $\hat{X}_M$, and Estimator II, $\hat{X}_F$, are feasible to compute. We now compare them with respect to both operational convenience and reliability.

In order to compute $\hat{X}_M$, the number of original households eligible for sampling from which the current residents come is needed. This information is particularly difficult to obtain at each successive wave of the survey. However, to compute $\hat{X}_F$ one only needs to know the number of current residents in the household (excluding new additions) and the number of residents from each original sample household. This information could be obtained from the 1979 Research Panel person identifier without collecting additional information.

The equal contribution from the members of a household is a natural assumption. It reflects better the actual share among the household members in the absence of knowledge of the actual contribution from each member. For example, without knowledge of each person's age, employment status and other needed information, it is more logical to assume that earnings and welfare benefits are equally contributed by household members than any arbitrary way of defining household members' shares. And as will be seen below, that heuristically $\hat{X}_F$ can be justified as the approximate minimum variance unbiased estimator under what seems to be natural assumptions given a state of ignorance about the actual shares of the household members.

Assume that at a subsequent wave time t three households are generated from two original households of the first wave of interview (time $t_o$) as follows:




Let $X_k(t_o)$, k=1, ..., $N(t_o)$ be the value of the characteristic of interest for household k at time $t_o$ and $X_j$, j=1, ..., N be that value for household j at time t. Using Section III we divide up $X_3$ in two parts, $fX_3$ and $(1-f)X_3$ and then associate $fX_3$ with $X_1$ and $(1-f)X_3$ with $X_2$. Without loss of generality, assume that a sample size of 1 was selected at the first wave, $t_o$, with probability $\pi_k$, k=1, ..., $N(t_o)$. An unbiased estimator of total, X, at time t can be written as

$$\hat{X} = \frac{\alpha_1}{\pi_1} X_1 + \frac{\alpha_2}{\pi_2} X_2 + \left( \frac{f\alpha_1}{\pi_1} + \frac{(1-f)\alpha_2}{\pi_2} \right) X_3 + \ldots$$

where $\alpha_i$, i=1, $N(t_o)$ is defined at the beginning of this section. Notice that both $\hat{X}_M$ and $\hat{X}_F$ are special cases of $\hat{X}$. The variance of $\hat{X}$ is

$$\mathrm{Var}(\hat{X}) = \pi_1 \left\{\left(\frac{1}{\pi_1} X_1 + \frac{f}{\pi_1} X_3\right) - X\right\}^2$$

$$+ \pi_2 \left\{\left(\frac{1}{\pi_2} X_2 + \frac{1-f}{\pi_2} X_3\right) - X\right\}^2 + \ldots\ldots$$

The remaining terms are not explicitly given here since they are not functions of f. The $\mathrm{Var}(\hat{X})$ is minimized if

$$f = \left( \frac{X_2+X_3}{\pi_2} - \frac{X_1}{\pi_1} \right) / \left( \frac{X_3}{\pi_1} + \frac{X_3}{\pi_2} \right)$$

Since usually not both $\pi_1$ and $\pi_2$ are known and in most of the surveys conducted by the Bureau of the Census, the inclusion probabilities, $\pi_k$, are about the same for all ultimate sampling units (even though they are unequal in the 1979 ISDP), one may simplify f to $f = (X_2 + X_3 - X_1)/2X_3$

Obviously, a weight defined as a function of survey observations is not easy to implement. To further simplify f, we assume the percentage growth of X from $t_o$ to t is constant for all units involved and define

$$a X_1(t_o) = X_1 + X_{31}$$

$$a X_2(t_o) = X_2 + X_{32}$$

$$X_3 = X_{31} + X_{32}$$

where $X_{3i}$ is the share of $X_3$ belonging to household members from original household i, i=1,2.

Without knowledge of both $X_1(t_o)$ and $X_2(t_o)$, one might naturally assume that the two initial households are about the same i.e., $X_1(t_o) = X_2(t_o)$ and reduce f to $X_{31}/(X_{31} + X_{32})$.

Now if the contribution to $X_3$ is proportional to the number of persons from each original household, then $f = S_{31}/S_{30}$, as defined in $W_{Fi}$. This result can be extended to any sample size as well as to the case that the new household members are from more than two original households. Therefore, without knowledge of the actual contribution from each household member, $\mathrm{Var}(\hat{X}_F)$ is smaller than $\mathrm{Var}(\hat{X}_M)$ under these assumptions.

## V. PROPOSED ESTIMATORS FOR LIST FRAMES

Since persons are the list frame sampling units, we can divide all persons in the general population into three groups based on their relationship with the list frame under consideration.

I) Persons who are included in the list frame (called list frame persons). For the SSI list frame, this group includes all the (under 65) recipients of the Federal Supplemental Security Income in December 1978; while for the BEOG list frame, this group includes all the eligible applicants of the Basic Educational Opportunity Grant as of September 1978 for school year 1978-79.

II) Persons who are not included in the list frame but live with a list frame person(s) during the first wave of interview (February through April 1979).

III) Persons who are not included in the list frame nor do they live with a list frame person(s) during the first wave of interview.

Both Group I and II had some chance to be included in the list frame sample, but Group III did not. The original (first quarter) households which consist of Group I and/or Group II persons will be called list frame households. As time went on, some members of Group III moved in and lived with person(s) belonging to Group I or II. Such members of Group III will be 'additions' for the list frame, since they are not initially eligible for sampling in the list frame. Note that the type of persons already described as "additions" for the general population (as defined in Section II) will also be "additions" for the list frame. For the following discussions, we now define two types of "additions" for the list frames: the "additions" that come from Group III will be called "Group III additions"and the type of "additions" as defined for the area frame will be called "area frame addition."

If a list of recipients of a government assistance program is used as a list frame then Group III is usually fairly large. If we construct our estimators the same way as we did for the area frame, we will include many of Group III persons in our estimates at time t of subsequent interviews. Consequently, we wouldn't really know what "subpopulation" we were estimating. In our opinion, it is not feasible to define such a subpopulation at time t. Without new sample drawn each wave from the updated list, a proper cross-sectional estimate for a list frame subpopulation at time t is not likely, especially if the turnover rate of the list frame members is high. Therefore, we will restrict our cross-sectional estimates to be based on only the original list frame sample persons (that is, the list frame persons selected for list frame sample plus all the persons who reside with them during the first quarter of interview) and the "area frame additions." In so doing we know that at any time t, the target population we are estimating consists of the original list frame subpopulation (that is Groups I and II) and the type of "additions" as defined in the area frame. Note that the original list frame subpopulation is determined by persons who were on the list at the time of sample selection. They may not be on the list by the time of initial interview.

In the 1979 ISDP panel, a household may have a multiple chance of being selected for the list frame sample if more than one member of the list frame persons live in that household at the first wave of interview. (Some effort was made to reduce multiple chance of selection for those households in SSI frame.) Therefore, the concept of the base weight for the first wave of interview is no longer trivial.

Similar to the area frame, we define $X(t_o) = \sum_{i=1}^{N^L(t_o)} X_i(t_o)$ as the population parameter to be estimated from a list frame sample at time $t_o$, where $X_i(t_o)$ is the value of the characteristic for the $i^{th}$ unit in the list frame subpopulation, which includes both Group I and II defined at the beginning of this section. Let

$\alpha_i$ = 1 if list frame person i is in the sample,
= 0 otherwise (note that $\alpha_i = 0$ for all non-list frame persons)

$\pi_i$ = the probability that list frame person i is selected for the list frame sample for the first wave of interview (time $t_o$) $= \Pr(\alpha_i = 1) = E(\alpha_i)$

$\beta_{oj}$ = the number of list frame persons (indexed by i) in the $j^{th}$ household at time $t_o$.

Then the base weight at time $t_o$ for the $j^{th}$ household and its residents is defined as

$$W_{oj} = \sum_{i=1}^{\beta_{oj}} \frac{\alpha_i}{\beta_{oj}\pi_i}$$

where $\alpha_i$ and $\pi_i$ are associated with list frame persons but are reindexed within household j.
For time t of a subsequent wave, let

$\beta_k$ = the total number of list frame persons living in the original (time $t_o$) list frame households which the current residents of the $k^{th}$ household come from.

$S_k$ = the total number of current residents at time t; $S_{k1}, S_{k2}, \ldots, S_{kr_k}$ be the number of current residents in the $k^{th}$ household who come from each of $r_k$ original list frame households; $S_{ka}$ is the number of current residents of the $k^{th}$ household who are from the "area frame additions"; and $S_{k\,III}$ is the number of current residents of the $k^{th}$ household who are from the "Group III additions." Therefore

$$S_k = \sum_{j=1}^{r_k} S_{kj} + S_{k\,III} + S_{ka} = S_{kc} + S_{ka}.$$

$N^L$ = the total number of units such as household or family units, in the list frame universe at time t

The two cross-sectional estimators for the total of a characteristic of the list frame target population at time t are as follows:

Estimator I (Multiplicity Estimator)

To avoid estimating "Group III additions" we will treat all the current residents from the "Group III additions" as a separate list frame sampling unit. Therefore, in the $k^{th}$ household at time t, there are $\beta_k + U_k$ list frame sampling units, where $U_k = 1$, if some of the current residents in the $k^{th}$ household are from "Group III additions," 0 otherwise. The multiplicity estimator for the list frame population total is given in the following:

$$\hat{X}_M^L = \sum_{k=1}^{N^L} W_{Mk}^L X_k$$

where

$$W_{Mk}^L = \sum_{i=1}^{\beta_k} \frac{\alpha_i}{(\beta_k + U_k)\pi_i}.$$

$\alpha_i$ and $\pi_i$ are associated with original list frame person but are reindexed within each current household k.

Estimator II (Fair Share Estimator)

Motivated by the assumption that all current residents contribute equally to a household we propose the following list frame estimator:

$$\hat{X}_F^L = \sum_{k=1}^{N^L} W_{Fk}^L X_k$$

where

$$W_{Fk}^L = \sum_{j=1}^{r_k} \frac{S_{kj}}{S_{kc}} W_{oj};$$

and $S_{kj}$ and $W_{oj}$ are associated with original household but are reindexed within each current household k.

These two estimators are constructed to estimate the list frame subpopulation excluding the "Group III addition." They are not unbiased estimators in the global sense, independent of the value of the characterisic of interest. However, the fair share estimator would be unbiased under the assumptions that all current residents contribute equally to a household and a household is treated as a fraction of a household if some of the current residents are from "Group III additions."

In addition to the "unbiasedness" described above, $\hat{X}_F^L$ is also preferred for the same reasons (operational and reliability) stated in the area frame. In computing $\hat{X}_F^L$, we need to know $\beta_{0j}$, the number of list frame persons in a sample household at the initial interview (time $t_0$). This information was not difficult to obtain. And at any subsequent wave of interview time t, we needed to know only $S_{kc}$, the total number of current residents who are not "area frame additions" and $S_{kj}$, the number of current residents from each original list frame sample household. The latter can be obtained through the person identifier.

However, in order to compute $\hat{X}_M^L$ at time t we would have to ask a complicated question to obtain $\beta_k$, the total number of list frame persons living in the original households from which the current residents come.

## VI. SUMMARY

These two estimators were constructed based on the specific procedure of following movers in the 1979 ISDP. However, they can be easily modified to apply to other designs and procedures. The fair share estimator was actually used for the 1979 ISDP. It is also being used for the 1984 Survey of Income and Program Participation.

As noted in Section III, the inverse of the inclusion probability of a household at time t is usually used as the weight of that household to obtain an unbiased estimator. When a household consists of members from two original households (called households i and j), the inclusion probability of this new household is $\pi_i + \pi_j - \pi_{ij}$, where $\pi_{ij}$ is the joint selection probability of households i and j at the first wave of interview. This inclusion probability is operationally impossible to obtain, but its inverse can be reduced to the weight ($W_{Mi}$) of the multiplicity estimator in most surveys conducted by the Census Bureau. In these surveys, the wave 1 inclusion probabilities are almost the same for all ultimate sampling units and the joint selection probabilities are generally due to the sampling without replacement within PSUs which are generally negligible. Therefore, the fair share estimator not only overcomes the trouble of obtaining such inclusion probabilities, but it has good variance properties under some reasonable conditions and it is easy to implement.

As described in Section V, the application of this approach to multiple frame longitudinal surveys presents additional problems, and the resulting estimators are not nearly as satisfactory.

This research was completed before the first interviews of the 1979 ISDP Research Panel. Horvitz and Folsom (1980) have done similar work in conjunction with the National Medical Care Utilization and Expenditure Survey.

## VII. ACKNOWLEDGMENTS

The author gratefully acknowledges the many helpful comments of Dr. D. Kasprzyk and other colleaques at the Census Bureau.

## REFERENCES

[1] David, M. (ed), (1983). Technical, Conceptual and Administrative Lessons of the Income Survey Development Program. New York: Social Science Research Council.

[2] Herriot, R. and Kasprzyk, D. (1984). The Survey of Income and Program Participation. To be published in the Proceedings of The Social Statistics Section, American Statistical Association.

[3] Horvitz, D.G. and Folsom, R.E. (1980). Methodological Issues in Medical Care Expenditure Surveys. The Proceedings of the Section on Survey Research Methods, American Statistical Association.

[4] Nelson, D., McMillen, D. and Kasprzyk, D. (June 1984). An Overview of the Survey of Income and Program Participation. SIPP Working Paper Series, Number 84-01.

[5] Survey Documentation for 1979 Research Panel, (1983). Department of Commerce, National Technical Information Service, Springfield Virginia.

[6] U.S. Department of Commerce, Bureau of the Census, (1978). Some initial thoughts on Weighting the 1979 Income Survey Development Program. Memorandum from G. Capps to M. Schooley, November 30, 1978.

[7] U.S. Department of Commerce, Bureau of the Census, (1979). Some Alternative Procedures on Weighting the 1979 Income Survey Development Program. Memorandum from M. Schooley to G. Shapiro, March 1, 1979.

[8] Whitmore, R., (ed.) (1982). Family Unit Weighting Methodology for the State Medicaid Household Survey Components of the National Medical Care Utilization and Expenditure Survey. Report Prepared for National Center for Health statistics and Health Care Financing Administration.

[9] Ycas, M. and Lininger, C., (1981). The Income Survey Development Program: Design Features and Initial Findings. Social Security Bulletin 44:11, November 1981.