Survey Methodology, June 1995
Vol. 21, No. 1, pp. 33–44
Statistics Canada

Weighting Schemes for Household Panel Surveys

GRAHAM KALTON and J. MICHAEL BRICK[1]

ABSTRACT

Household panel surveys often start with a sample of households and then attempt to follow all the members of those households for the life of the panel. At subsequent waves data are collected for the original sample members and for all the persons who are living with the sample members at the time. It is desirable to include the data collected both for the original sample persons and for the persons living with them in making person-level cross-sectional estimates for a particular wave. Similarly, it is desirable to include data for all the households for which data are collected at a particular wave in making household-level cross-sectional estimates for that wave. This paper reviews weighting schemes that can be used for these purposes. These weighting schemes may also be used in other settings in which units have more than one way of being selected for the sample.

KEY WORDS: Cross-sectional estimates; Fair share weighting; Multiplicity weighting; Panel surveys; Weight share method.

1. INTRODUCTION

National panel surveys of household economics have been mounted in many countries in recent years. The U.S. Panel Study of Income Dynamics (PSID), conducted by the Survey Research Center of the University of Michigan, began in 1968 and has been collecting data on an annual basis since that time (Hill 1992), and the British Household Panel Survey began in 1990 (Buck *et al.* 1994). Similar household panel surveys are also in progress or are being planned in most other European countries. The U.S. Bureau of the Census started to conduct the Survey of Income and Program Participation (SIPP) in 1983 (Nelson *et al.* 1985; Kasprzyk 1988; Jabine *et al.* 1990; Citro and Kalton 1993), and Statistics Canada introduced the Survey of Labour and Income Dynamics (SLID) in 1994 (Lavallée *et al.* 1993).

A common feature to most of these household panel surveys is that they start with a national sample of households, and then follow all the members of those households for the life of the panel. Over the course of time, household compositions change in a variety of ways. Some members of original sampled households leave those households to set up on their own or to join other households, as, for example, when a daughter leaves her parental household to get married. New members may join original sampled households, as, for example, when an elderly parent moves in with the family of a child or when a child is born to a household member. In order to be able to describe the economic circumstances of sample members at different points of time, household panel surveys usually collect data not only for the sample members but also for the individuals living with the sample members at the particular point of time. Following Lavallée (1995), these individuals are termed cohabitants in this paper. In other literature, they are often called associated persons or nonsample persons.

As the panel duration increases, the proportion of cohabitants in the sample at a wave rises. For example, in the 1984 SIPP panel, cohabitants comprise about 8.6 percent of the sample after one year and about 12.6 percent of the sample after two years (based on Table 1 in Kasprzyk and McMillen 1987). With a long-term household panel survey, the proportion of cohabitants becomes substantial after several years. The PSID, for example, defines sample members as all persons in the family units sampled in 1968 who are still alive, all the children born to these original sample members since the start of the panel, and the children of such children. In addition, the PSID collects data on the cohabitants who are living with sample members at each individual wave of data collection. Of the 20,535 individuals in interviewed family units in 1992, 41.2 percent were original sample members, 34.6 percent were the children of original sample members born since the start of the panel and children of such children, and 24.2 percent were cohabitants (excluding the Latino sample that was added in 1990) (Hill 1995).

This paper reviews methods of weighting the data collected from both sample persons and cohabitants in order to produce unbiased (or approximately unbiased) estimates of population parameters. In considering the analysis of a household panel survey, three different types of analysis may usefully be distinguished:

[1] Graham Kalton and J. Michael Brick, Westat Inc., 1650 Research Blvd., Rockville, MD 20850, U.S.A.

- Cross-sectional analyses of households at a particular point in time;
- Cross-sectional analyses of individuals at a particular point in time;
- Longitudinal analyses of individuals over a period of time.

Weighting schemes for these three types of analysis are discussed in later sections. Longitudinal analyses of households over a period of time are not treated here because of the problematic nature of this type of analysis caused by changes in household composition (see, for example, Duncan and Hill 1985).

The weighting schemes used in household panel surveys need to account for the fact that households and individuals included in the survey at a particular wave may have more than one route by which they can be selected. At a given wave a household and its members are included in the sample if any of the original households (*i.e.*, households existing at the time of the initial selection) from which the current household has drawn members was selected. With the usual weighting approach, households are assigned weights inversely proportional to their joint selection probabilities, taking account of the different ways they can be selected. However, this approach cannot be applied with most household panel surveys because these joint selection probabilities cannot be determined. The alternative weighting approach reviewed here, termed by Lavallée (1995) the *weight share method*, avoids the need to know the joint selection probabilities of sample elements, but it introduces a random variation into the weights. Since this random variation results in a loss in precision of the survey estimates as compared with the inverse selection probability weighting scheme, this alternative approach should be considered only for situations where the joint selection probabilities cannot be ascertained. This situation often applies in household panel surveys and also in a number of other sample designs where elements can be selected by different routes.

In order to prepare for the discussion of weighting schemes for household panel surveys, the next section elaborates on the household changes that can occur over time, and the types of individuals involved. Sections 3, 4 and 5 then discuss weighting schemes that may be used for the three different forms of analysis described above. These sections deal with weighting schemes for unequal selection probabilities, without the complications of adjustments for nonresponse and noncoverage. The discussion relies heavily on previous work by Ernst (1989), Gailly and Lavallée (1993), Huang (1984), Judkins *et al.* (1984), Lavallée and Hunter (1992), and Little (1989). Section 6 then briefly reviews the issues involved in making adjustments to the weights to compensate for missing data arising from nonresponse and noncoverage. Section 7 presents some concluding remarks, and provides an illustration of another application of the weight share method.

2. CHANGES IN POPULATION AND HOUSEHOLD COMPOSITION OVER TIME

In analyzing a panel survey, it needs to be recognized that survey populations change over time. With household panel surveys it is important to distinguish between changes in population composition and changes in household composition.

The composition of a survey population changes over time because some individuals leave the population, some enter the population, and some may leave and join the population more than once. Individuals leave the population through death, emigration, or entering an institution (for surveys of the noninstitutional population). They enter the population through birth (or reaching the specified minimum age), immigration, and leaving an institution.

Households change composition over time for many different reasons, including deaths, births, marriages and divorces. For example, a household at time 1 may contain several individuals who end up in a number of different households at time 2. These individuals may set up new households on their own, they may join individuals who were in one or more households at time 1, or they may join individuals who were not in the population at time 1. One or more of the individuals may leave the population during the intervening period.

Consider a simple sample design in which households are selected independently at time 1 with equal probability. At time 2, the sample of households comprises all the households that contain one or more individuals from the households sampled at time 1, and the sample of individuals at time 2 comprises all the members of the sampled households at time 2. The samples of households and individuals at time 2 are selected with unequal probabilities. For instance, the selection probability of a household at time 2 that contains individuals from three households at time 1 is three times greater than that of a household at time 2 that contains individuals from only one household at time 1. Similarly, the individuals in that household have three times the probability of selection. Thus weighting schemes that compensate for these unequal selection probabilities are needed for the analysis of the resultant data.

Changes in population composition occur when individuals leave or enter the population. An individual sampled at time 1 who leaves the population before time 2 reduces the sample size for time 2 but does not otherwise affect cross-sectional estimates at time 2. In essence, the sampling frame for the time 2 population is the time 1 population, with the leavers in the intervening period being treated as blanks on the frame. Simply omitting the selected blanks from the time 2 sample causes no bias in the survey estimates (see, for example, Kish 1965). The situation with regard to entrants is, however, less straightforward. The household panel survey enumeration rule described above incorporates new entrants who join households that contain individuals who were eligible for the initial sample into the population for cross-sectional estimates for later time points. However, new entrants who set up their own households are not represented in person-level analyses at later waves of the panel. Equally, households composed of only new entrants are not represented in household-level analyses at later waves.

The failure of household panel surveys to cover households composed of only new entrants presents a problem for cross-sectional analyses of later waves of the panel. If these households and their members constitute a negligible proportion of the population, the solution may be to simply ignore the problem. However, if the proportion is appreciable, as can occur in later waves of a long-term panel, alternative solutions may be called for. One possibility is to add a supplementary sample of new entrants (*e.g.*, immigrants) to the panel, as discussed by Lavallée (1995) for the SLID. This solution is, however, often impracticable. Another solution is to limit the population of inference to persons who were members of the population at the start of the panel. New entrants found living with sample members are then excluded from the sample. This solution provides a clearcut definition of the population of inference. Whether the solution is appropriate depends on whether that definition can adequately satisfy the survey objectives.

Changes in population composition pose problems for longitudinal analyses of individuals. For many purposes, the population of inference is restricted to those who were present in the population throughout the time period of observation specified for the analysis. The inclusion of cohabitants in longitudinal analysis also creates problems. If the time period for the longitudinal analysis starts at the beginning of the panel, the analysis can be restricted straightforwardly to original sample members. If the time period starts later, it is tempting to include both original sample members and cohabitants joining the panel before the start of the analytic time period. However, the usual enumeration rules for household panel surveys specify that data are collected for cohabitants only while they continue to live with original sample members, that is, they are not followed if they cease to live with such persons. Unless the time period is short enough that the number of cohabitants who cease to live with sample persons in that period is negligible, this enumeration rule makes it problematic to include cohabitants in longitudinal analyses. This problem is discussed further in Section 5.

3. CROSS-SECTIONAL ESTIMATES FOR HOUSEHOLDS

This section considers weighting schemes that may be used to produce cross-sectional estimates for households for any wave of a household panel survey after the first. At the first wave a sample of households is selected and all the individuals in the sampled households become panel members to be followed throughout the life of the panel or until they leave the survey population. At a subsequent wave, wave t, the household sample comprises all the households in which panel members reside. Households that consist of new entrants only are not represented in the sample at later waves. Such households are ignored here. Complications of nonresponse are deferred until Section 6.

Consider the estimation of the total Y for all H households in the population at time t:

$$Y = \sum_{i=1}^{H} Y_i. \tag{3.1}$$

A general estimator for this total can be expressed as

$$\hat{Y} = \sum_{i=1}^{H} w_i Y_i,$$

where w_i is a random variable that takes the value $w_i = 0$ if household i is not in the sample. The expectation of $\hat{Y}$ is

$$E(\hat{Y}) = \sum_{i=1}^{H} E(w_i) Y_i. \tag{3.2}$$

By comparing equations (3.1) and (3.2), it can be seen that $\hat{Y}$ is unbiased for Y for any weighting scheme for which $E(w_i) = 1$ for all i.

There are many ways to satisfy the condition $E(w_i) = 1$. Three will be treated here. First, consider a standard *inverse selection probability weighting scheme*. The probability of a household being in the sample at time t is the probability of one or more of the households at time 1 from which it has drawn members being selected for the original sample. The probability of household H_i being in the sample at time t is then

$$\begin{aligned} P(H_i) &= P(h_j \cup h_k \cup h_\ell \cup \ldots) \\ &= \sum p_j - \sum\sum p_{jk} + \sum\sum\sum p_{jk\ell} - \ldots, \end{aligned} \tag{3.3}$$

where $P(h_j \cup h_k \cup h_\ell \cup \ldots)$ is the selection probability of the union of original households h_j, h_k, h_ℓ, *etc.* for the original sample, p_j is the selection probability of original household h_j for the original sample, p_{jk} is the joint selection probability of original households h_j and h_k for the original sample, *etc.* and where households h_j, h_k, h_ℓ, *etc.* each contain at least one member who is currently in household H_i. The weight for each sampled household is then $w_i = 1/P(H_i)$. With this weighting scheme,

$$E(w_i) = P(H_i)[1/P(H_i)] + [1 - P(H_i)]0 = 1,$$

satisfying the condition for an unbiased estimator of a population total.

In practice, the computation $P(H_i)$ will generally not be as complex as equation (3.3) might suggest because the number of original households represented in household H_i is usually small. With, say, two original households involved, $P(H_i)$ reduces to

$$P(H_i) = P(h_1 \cup h_2) = p_1 + p_2 - p_{12}. \quad (3.4)$$

A problem with the application of the inverse selection probability approach is that p_j may be known only for households selected for the original sample, and not for other households. Also the joint probability may not be known. Even when the original sample was selected with equal probabilities, so that all the p_j are the same, the joint probability may depend on the sample design (for instance, whether the two households were in the same segment or not). The difficulty of obtaining $P(H_i)$ is a major drawback with the inverse selection probability approach.

An alternative strategy for developing the weights for time t is to base them only on the selection probabilities of households selected for the original sample, thus avoiding the difficulty in obtaining $P(H_i)$ noted above. One approach is to identify the set of households h_j at time 1 that would result in household H_i being in the sample at time t, and compute the weight for household H_i as

$$w_i = \sum_j \alpha_{ij} w'_{ij}, \quad (3.5)$$

where $w'_{ij} = 1/p_j$ if household h_j, which has at least one member in household H_i, was selected for the original sample and $w'_{ij} = 0$ if not, and where α_{ij} are any set of constants satisfying $\sum_j \alpha_{ij} = 1$.

With this approach,

$$E(w'_{ij}) = p_j(1/p_j) + (1 - p_j)0 = 1,$$

and hence

$$E(w_i) = \sum_j \alpha_{ij} = 1.$$

Thus, the use of weights w_i will yield unbiased estimators of totals for the household population for any choice of constants α_{ij}, provided that $\sum_j \alpha_{ij} = 1$. As indicated above, the principal advantage of this type of scheme is that it requires information only on the initial selection probabilities of the original households that were sampled at time 1, which are known. It does not require information on the initial selection probabilities of the other original households that have members in the current household, which are often not known.

A natural choice of α_{ij} is to make them equal for all the original households that lead to the selection of household H_i at time t. Huang (1984) terms this scheme a multiplicity approach. Here the scheme will be called an *equal household weighting scheme*. With this scheme

$$w_i = \sum_j w'_{ij}/C_i, \quad (3.6)$$

where C_i is the number of original households represented in household H_i at time t.

An alternative version of the above approach is one based on original sample persons rather than households. In this case, let I_{ijk} denote individual k from original household j in household i. Then

$$w_i = \sum_j \sum_k \alpha_{ijk} w'_{ijk},$$

where $w'_{ijk} = 1/p_j$ if individual k in household h_j was in the original sample and $w'_{ijk} = 0$ if not, and where the α_{ijk} are any set of constants satisfying $\sum_j \sum_k \alpha_{ijk} = 1$. Since the probability of an individual being selected for the original sample is the same as that of that individual's household,

$$E(w'_{ijk}) = p_j(1/p_j) + (1 - p_j)0 = 1.$$

In this case, the natural choice of the constants α_{ijk} is to make them equal for all members of the current household who were eligible for selection for the original sample. This produces what has been termed the fair share weighting scheme (Huang 1984; Ernst 1989). This scheme is termed here an *equal person weighting scheme*. With this scheme

$$w_i = \frac{1}{M_i} \sum_j M_{ij} w'_{ij},$$

where $w'_{ij} = w'_{ijk}$ is constant for all individuals in household H_i emanating from the same sampled household at time 1, M_{ij} is the number of individuals in household H_i coming from household h_j, and $M_i = \sum_i M_{ij}$ is the number of individuals in household H_i who were eligible for the sample at time 1. The equal person weighting scheme is applied in the SIPP and is proposed for use in the SLID.

Although developed here in terms of persons rather than households, it is readily apparent that the equal person weighting scheme could equally have been generated in terms of households. As shown above, the household weight $w_i = \sum_j \alpha_{ij} w'_{ij}$ satisfies the condition $E(w_i) = 1$ for any set of constants α_{ij} such that $\sum_j \alpha_{ij} = 1$. The equal household weighting scheme chooses $\alpha_{ij} = 1/C_i$, with $\sum_j \alpha_{ij} = 1$. The choice $\alpha_{ij} = M_{ij}/M_i$, with $\sum_j \alpha_{ij} = 1$, leads to the equal person weighting scheme.

It is instructive to compare the inverse selection probability weighting scheme with the equal household and equal person weighting schemes in a simple case. Following Little (1989), consider household H_i selected at time t

with household members coming from two original households. Let p_1 and p_2 denote the selection probabilities for the original households, and let p_{12} denote their joint selection probability. Under the inverse selection probability approach, the household weight is

$$w_i^* = \frac{1}{p_1 + p_2 - p_{12}},$$

as indicated above.

Under the equal person weighting scheme the weight for household H_i depends on which household or households were selected for the original sample:

$w_i = P_1/p_1$ if only household h_1 was selected;

$w_i = P_2/p_2$ if only household h_2 was selected;

$w_i = (P_1/p_1) + (P_2/p_2)$ if both h_1 and h_2 were selected;

where P_1 and P_2 are the proportions of members of household H_i who came from households h_1 and h_2, respectively (excluding any new entrants to the population). The probability of only household h_1 being selected is $(p_1 - p_{12})$, of only household h_2 being selected is $(p_2 - p_{12})$, and of both households being selected is p_{12}. The expected value of the weight conditional on household H_i being in the sample is thus

$$E(w_i \mid H_i \text{ in sample}) =$$

$$\frac{(p_1-p_{12})(P_1/p_1)+(p_2-p_{12})(P_2/p_2)+p_{12}[(P_1/p_1)+(P_2/p_2)]}{p_1+p_2-p_{12}},$$

i.e.,

$$E(w_i \mid H_i \text{ in sample}) = \frac{1}{p_1 + p_2 - p_{12}} = w_i^*.$$

As this result demonstrates, the weight for household H_i varies depending on which original households were selected, but in expectation the weight is the same as that obtained from the inverse selection probability approach.

Results for the expectation of the weight of household H_i under the equal household weighting scheme can be readily obtained as a special case of the above derivation in which $P_1 = P_2 = ½$. In expectation, the weight is the same as that for the inverse selection probability approach.

Given that the weight $w_i = \sum_j \alpha_{ij} w'_{ij}$ satisfies the condition $E(w_i) = 1$ for any set of α_{ij} such that $\sum_j \alpha_{ij} = 1$, the question arises as to the optimal choice of the α_{ij}. One approach is to choose the α_{ij} to minimize the variance of the estimated total $\hat{Y}$.

The variance of $\hat{Y}$ may be expressed as

$$V(\hat{Y}) = VE(\hat{Y} \mid s) + EV(\hat{Y} \mid s), \qquad (3.7)$$

where s denotes the set of households in the sample at time t. Now

$$E(\hat{Y} \mid s) = E\left(\sum_{i=1}^{H} w_i Y_i \mid s\right)$$

$$= \sum^{s} E(w_i \mid H_i) Y_i = \sum^{s} w_i^* Y_i = \hat{Y}^*,$$

where $\hat{Y}^*$ is the standard inverse selection probability estimator. Thus

$$VE(\hat{Y} \mid s) = V(\hat{Y}^*).$$

The first term in equation (3.7) is thus the variance of the standard inverse selection probability estimator, and the second term is the additional variance resulting from the use of weighting schemes from the class (3.5), $w_i = \sum_j \alpha_{ij} w'_{ij}$. The α_{ij} may then be chosen to minimize $EV(\hat{Y} \mid s)$.

Consider

$$V(\hat{Y} \mid s) = V\left(\sum^{H} w_i Y_i \mid s\right)$$

$$= \sum^{s} Y_i^2 V(w_i \mid H_i) +$$

$$\sum\sum_{i \neq i'} Y_i Y_{i'} \mathrm{Cov}(w_i, w_{i'} \mid H_i, H_{i'}).$$

Assuming $\mathrm{Cov}(w_i, w_{i'} \mid H_i, H_{i'}) = 0$,

$$V(\hat{Y} \mid s) = \sum Y_i^2 V(w_i \mid H_i)$$

$$= \sum Y_i^2 [E(w_i^2 \mid H_i) - w_i^{*2}],$$

since, as noted above, $E(w_i \mid H_i) = w_i^*$. Thus, assuming $\mathrm{Cov}(w_i, w_{i'} \mid H_i, H_{i'}) = 0$, $V(\hat{Y} \mid s)$ is minimized when $E(w_i^2 \mid H_i)$ is minimized.

Consider the application of this approach to the simple case discussed above in which H_i is composed of members from two original households and let $w_i = \alpha_i w'_{i1} + (1 - \alpha_i) w'_{i2}$. Then

$$E(w_i^2 \mid H_i) =$$

$$\frac{(p_1-p_{12})\dfrac{\alpha_i^2}{p_1^2} + (p_2-p_{12})\dfrac{(1-\alpha_i)^2}{p_2^2} + p_{12}\left(\dfrac{\alpha_i}{p_1} + \dfrac{1-\alpha_i}{p_2}\right)^2}{p_1+p_2-p_{12}}.$$

Minimizing $E(w_i^2 \mid H_i)$ is equivalent to minimizing

$$\Delta = (p_1 - p_{12}) p_2^2 \alpha_i^2 + (p_2 - p_{12}) p_1^2 (1 - \alpha_i)^2$$

$$+ p_{12}[(p_2 - p_1)\alpha_i + p_1]^2.$$

Then

$$\frac{\partial \Delta}{\partial \alpha_i} = 2(p_1 - p_{12})p_2^2\alpha_i - 2(p_2 - p_{12})p_1^2(1 - \alpha_i) + 2p_{12}(p_2 - p_1)[(p_2 - p_1)\alpha_i + p_1].$$

Solving $\partial\Delta/\partial\alpha_i = 0$ for α_i gives the optimum α_i as

$$\alpha_{oi} = \left(1 + \frac{p_2 - p_{12}}{p_1 - p_{12}}\right)^{-1}. \quad (3.8)$$

If the original households are selected independently, *i.e.*, $p_{12} = p_1 p_2$,

$$\alpha_{oi} = \left[1 + \frac{p_2(1 - p_1)}{p_1(1 - p_2)}\right]^{-1} = \left[1 + \frac{\psi_2}{\psi_1}\right]^{-1}, \quad (3.9)$$

where $\psi_j = p_j/(1 - p_j)$ is the odds of original household h_j being selected.

Irrespective of whether the households are sampled independently, in the special case of an equal probability (epsem) sample of households initially, with $p_1 = p_2$,

$$\alpha_{oi} = \frac{1}{2}.$$

Thus, in the two-household case, the equal household weighting scheme minimizes the variance of the household weights around the inverse selection probability weight when the initial sample is an epsem one.

The optimal choice of α_{oi} given by (3.8) requires knowledge of p_1, p_2 and p_{12}, and that given by (3.9) requires independence and knowledge of p_1 and p_2. If these probabilities were known, then the standard inverse selection probability weight could be employed and would be preferable. In the case of an approximately epsem sample, the equal household weighting scheme should be close to the optimal, at least for the case where the members of the household at time t come from one or two households at the initial wave. This would apply, for instance, in the case of an epsem initial sample, with perhaps a few departures from epsem. With the equal household weighting scheme, when only one of the C_i original households, h_j, represented in H_i was selected for the original sample (as will generally be the case), then the weight for H_i is simply $1/C_i p_j$.

In the case of a non-epsem initial sample, the choice of the α_{ij} would ideally depend on the original household selection probabilities. However, since these probabilities are unknown, that approach cannot be applied. By default, the equal household or equal person weighting schemes may therefore be employed in this case. The use of these schemes (or any scheme with constant α_{ij}'s satisfying $\sum_j \alpha_{ij} = 1$) with a non-epsem initial sample still results in an unbiased estimate $\hat{Y}$. The drawback to these schemes in such a case is only that the α_{ij} are suboptimal in terms of minimizing the variance of $\hat{Y}$.

It should be noted that the equal household weighting scheme requires information on the number of original households h_j contributing members to household H_i at time t. That number may be difficult to determine in some cases. Consider, for example, a household at time t that contains two cohabitants. It may sometimes be difficult to determine whether these two persons were in a single household or in two separate households at the time of the initial sample selection. The equal person weighting scheme has the attractive feature of avoiding the need for Wave 1 household information, except for persons in sampled households at Wave 1. This feature provides an important reason for preferring the equal person to the equal household weighting scheme.

4. CROSS-SECTIONAL ESTIMATES FOR INDIVIDUALS

In producing cross-sectional estimates for individuals for any wave of a household panel survey after the first, it needs to be recognized that some new entrants will have joined the survey population since the start of the panel. New entrants who join households that contain one or more members of the original population can be represented in cross-sectional estimates for later waves, but new entrants living in households that do not contain any members of the original population are not covered (unless a special sample of them can be taken). The former type of new entrants is included in the weighting procedure described below, but the latter type is not.

Let there be N individuals in the population at time t, with N_i individuals in household H_i $(i = 1, 2, \ldots, H)$ and $\sum N_i = N$. The members of household H_i come from households h_j, h_k, h_ℓ, *etc.*, at time 1. Let M_{ij} denote the number of members of household H_i at time t who were in household h_j at the start of the panel. The sum $M = \sum\sum M_{ij}$ is less than the population size at time 1 because of leavers from the population in the period from time 1 to time t, and $M < N$ because of new entrants to the population who are in households containing members from the original population.

Consider now the estimation of a total for the population of individuals at time t:

$$Y = \sum_{i=1}^{H} \sum_{k=1}^{N_i} Y_{ik}. \quad (4.1)$$

where Y_{ik} is the value for individual k in household H_i. As in the household case discussed in the previous section, a general estimator for this total can be expressed as

$$\hat{Y} = \sum_{i=1}^{H} \sum_{k=1}^{N_i} w_{ik} Y_{ik}, \tag{4.2}$$

where w_{ik} is a random variable that takes the value $w_{ik} = 0$ if individual k in household H_i is not in the sample. The estimator $\hat{Y}$ is unbiased for Y provided that $E(w_{ik}) = 1$ for all i and k.

As noted earlier, there are many ways to satisfy the condition $E(w_{ik}) = 1$. It is instructive to consider three of them. First, let $w_{ik} = 0$ for all individuals not in the original sample. In this case, the estimator $\hat{Y}$ discards cohabitants. Let p_{ik} denote the probability of a member of the original population, individual k residing in household H_i at time t, being selected for the initial sample, and let $w_{ik} = 1/p_{ik}$. Then, for such an individual

$$E(w_{ik}) = p_{ik}(1/p_{ik}) + (1 - p_{ik})0 = 1.$$

With this scheme, all new entrants to the population have $w_{ik} = 0$ with certainty. Thus $\hat{Y}$ in (4.2) provides an unbiased estimator of the total for the original population that is still present at time t, but does not include a component for the new entrants.

Modifications to the above procedure can be made to cover certain types of new entrants. For instance, births to sampled mothers can be included by assigning them the weights of their mothers, or if, as in the SIPP, the survey population is taken to be adults aged 16 and over, those under 16 at the start of the panel can be treated as sampled persons with assigned probabilities, and they can be included in the analyses of later waves after they have attained the age of 16. Such modifications do not, however, handle all types of new entrants. Provided that the proportion of other types of new entrants is small, this deficiency may not be a serious concern.

The weighting scheme that restricts the analysis to original sample persons, plus certain specified new entrants, is employed with the PSID. Its limitation is that it fails to make direct use of data collected for cohabitants. Such data may be used to provide information on the situation of sample persons, but the cohabitants are excluded from the sample for the analysis.

In order to include cohabitants in cross-sectional analyses for time t they need to be assigned positive weights. Noting that the probability of an individual being selected for the sample is the same as that of his or her household, weighting schemes for cross-sectional analyses of individuals at wave t can be obtained directly from those for households given in Section 3. Here we will develop the general strategy of producing weights for cross-sectional analysis at time t based only on the selection probabilities of members of the original sample, thus avoiding the problems with the inverse selection probability approach noted in Section 3.

Let I_{ijk} denote individual k from original household h_j who is now in household H_i. Let w_i denote the weight for every member of household H_i for cross-sectional analyses at time t, and let

$$w_i = \sum_j \sum_k \alpha_{ijk} w'_{ijk}$$

where $w'_{ijk} = 1/p_j$ if household h_j was in the original sample and $w'_{ijk} = 0$ if not. Then, as before, $E(w'_{ijk}) = 1$ for members of the original population. New entrants, for whom $p_j = 0$, may be handled by setting $\alpha_{ijk} = 0$. Then

$$E(w_i) = \sum_j \sum_k^{N_i} \alpha_{ijk} E(w'_{ijk}) = \sum_j \sum_k^{M_i} \alpha_{ijk} = 1$$

provided that $\sum_j \sum_k \alpha_{ijk} = 1$. Under this condition $\hat{Y}$ is unbiased for Y.

A natural choice of α_{ijk} is to set $\alpha_{ijk} = 1/M_i$ for all members of the original population. This is the equal person weighting scheme in which every member of household H_i at time t (including new entrants) receives the weight

$$w_i = \sum_j \sum_k w'_{ijk}/M_i.$$

Another choice of the α_{ijk} is that used for the equal household weighting scheme. Let C_i denote the number of original households that have members in household H_i at time t. Then $\sum_j \sum_k \alpha_{ijk} = 1$ can be divided equally between households, with each member of original household h_j being assigned a value of $\alpha_{ijk} = 1/C_i M_{ij}$. Then for original household h_j

$$\sum_k^{M_{ij}} \alpha_{ijk} = 1/C_i.$$

The derivation of the α_{ijk} to minimize the variance of the estimated total $\hat{Y}$ for the population of individuals follows directly from the corresponding derivation for the population of households given in Section 3. The estimated total for the population of individuals is

$$\hat{Y} = \sum_i^{s} \sum_k^{N_i} w_{ik} Y_{ik} = \sum_i^{s} \sum_k^{N_i} w_i Y_{ik},$$

since the weights for every individual in sampled household H_i are the same. This estimated total can be expressed as

$$\hat{Y} = \sum_i^{s} w_i Y_i,$$

where $Y_i = \sum_k Y_{ik}$ is the household total for H_i. Thus $\hat{Y}$ can be expressed as a household total, and the results of Section 3 can be applied directly.

Consider the example from Section 3 in which H_i is composed of members from only two original households, perhaps together with one or more new entrants. In this case the person-level weight $w_i = \sum_j \sum_k \alpha_{ijk} w'_{ijk}$ reduces to

$$w_i = \left(\sum_k \alpha_{i1k} \right) w'_{i1} + \left(\sum_k \alpha_{i2k} \right) w'_{i2}$$

$$= \alpha_i w'_{i1} + (1 - \alpha_i) w'_{i2},$$

where $\alpha_i = \sum_k \alpha_{i1k}$. As shown in equation (3.8), the optimum value of α_i is

$$\alpha_{oi} = \left(1 + \frac{p_2 - p_{12}}{p_1 - p_{12}} \right)^{-1}.$$

The individual values α_{ijk} are not needed for computing the w_i; only the original household totals $\sum_k \alpha_{ijk}$ are required. If individual values are needed for the α_{ijk}, they may be simply assigned as $\sum_k \alpha_{ijk} / M_{ij}$.

As in the household case, the optimum weighting α_{oi} requires knowledge of p_1, p_2 and p_{12}. If these probabilities are known, the standard inverse selection probability weight w_i^* can be computed, and would be preferred. In the case of an approximately epsem sample, the equal household weighting scheme should fare well. However, the equal household weighting scheme requires information on the number of original households contributing members to current household H_i, and this information may not always be available. As discussed in Section 3, for this reason the equal person weighting scheme may be preferred.

5. LONGITUDINAL ANALYSES OF INDIVIDUALS

A key analytic advantage of a panel survey is the ability to conduct longitudinal analyses relating variables for the same sampled units measured at different time points. Since all persons in original sampled households are followed throughout the life of the panel or until they leave the survey population, the data they provide may be readily analyzed longitudinally for any time period within the panel's time span (although nonresponse adjustments may be needed for panel attrition). Thus, for example, in a ten-year panel, data for original sampled persons may be analyzed from year 1 to year 10, from year 5 to year 9, or for any other period. New entrants (*e.g.*, births) may be included in the analysis for periods beginning after the start of the panel provided that they are treated as panel members who are followed throughout the panel even when they leave the households of original sampled persons.

Given the weighting schemes described in the previous section, cohabitants can be included in cross-sectional analyses of later waves. These weighting schemes provide a cross-sectional representation of the population at any wave of the panel (apart from new entrants not living with original population members). It is then possible to consider all the sample of original sample members and cohabitants at time t as the initial sample of a new panel that may be used for longitudinal analyses from time t to $(t + k)$. This procedure is, for instance, used in the SIPP, where all original sample members and cohabitants present at the start of the second year of the panel are included in analyses relating to that year.

The limitation to the inclusion of cohabitants in longitudinal analysis is that the following rules used in most household panel surveys specify that cohabitants are dropped from the panel if they cease living with original sample persons. Thus, cohabitants who live with original sample members at the start of the analysis period but who cease to live with them before the end of that period effectively become nonrespondents. If the analysis period is relatively short, the number of such nonrespondents may be small and the risk of serious nonresponse bias may be negligible. If the analysis period is a long one, however, the number of not-followed cohabitants may be appreciable, causing concerns about potential bias. The issue here is one of a trade-off between the reduced variance due to the increase in sample size from including cohabitants in the analysis versus the increased bias resulting from the additional nonresponse caused by failing to follow cohabitants leaving the households of original sample persons.

The additional nonresponse bias can be avoided by changing the following rules to specify that cohabitants are to be followed from the time they join the panel for the rest of the life of the panel, or until they leave the survey population, irrespective of whether they continue to live with original sample members. This change, however, leads to an expanding panel and the need for additional resources. Not only do data need to be collected for cohabitants at waves after they cease to live with sample persons, but data also need to be collected for any persons with whom the cohabitants live at later waves.

6. ADJUSTMENTS TO COMPENSATE FOR NONRESPONSE AND NONCOVERAGE

The discussion thus far has assumed that data are collected for all sampled persons and their cohabitants and that all the target population is covered by the sampling procedures. In practice both these assumptions are violated. Nonresponse is present in nearly all surveys and is of particular concern in household panel surveys, where some sampled households fail to respond at the initial wave and others fail to respond at some of the subsequent waves. The sampling frames used in most surveys are subject to some degree of noncoverage, and in later waves of household panel surveys there is an additional source of noncoverage associated with new entrants to the population who are not living with members of the original population.

In a simple cross-sectional survey, missing data can be classified into item nonresponse, total nonresponse and noncoverage. Imputation procedures can then be used to assign values for item nonresponses, weighting adjustments can be applied to compensate for total nonresponse, and poststratification adjustments can be applied to compensate for nonresponse and noncoverage. The situation is made far more complex in panel surveys by the occurrence of wave nonresponse, which arises when a sampled element responds for some but not all of the waves for which it was eligible. Not only do methods need to be devised to compensate for wave nonresponse, but also the preferred methods of compensation may depend on the type of analysis to be performed, in particular whether cross-sectional or longitudinal analyses are to be conducted.

From one perspective wave nonresponse can be viewed as a set of item nonresponses in the element's longitudinal record, suggesting that imputation may be used to fill in the missing values. Alternatively, it can be treated as total nonresponse, handled by weighting adjustments. The imputation approach is more natural for the creation of a panel file for longitudinal analysis, whereas the weighting approach is more natural for the creation of a cross-sectional file for the analysis of the data collected at a single wave.

The attraction of the imputation approach with a longitudinal file is that it retains all the reported data, whereas the weighting approach discards the reported data for all the elements that fail to provide data for one or more waves for which they were eligible. However, the imputation approach may involve the fabrication of a large amount of data, especially when an element fails to respond at several waves. Thus, for panel files, a compromise solution may be preferred, imputing responses for elements with few missing waves and using weighting adjustments to compensate for those with several missing waves (including total nonrespondents). In the SIPP, for example, imputation is used to assign responses for sample persons with a single missing wave that is bounded on both sides by responding waves, and weighting adjustments are used for all other sample persons with missing waves (Singh *et al.* 1990). Further discussion of methods of handling wave nonresponse in panel files is provided by Lepkowski (1989), Kalton (1986), and Lepkowski *et al.* (1993).

Another complication of some household panel surveys is the occurrence of partial household nonresponse, which occurs when the survey data are collected for some but not all members of a sampled household at a particular wave. The lack of data for one individual in a household means that key household characteristics (*e.g.*, household earnings) cannot be computed. One solution is to drop the household and its responding members from the sample, and use a weighting adjustment. Another is to impute the responses for the nonresponding household members, as is done in SIPP (where they are termed Type z nonrespondents). With the latter solution, data are available for all members of responding households, and hence person-level adjustments are unnecessary within responding households.

We now turn to consider the issues involved in dealing with missing data for cross-sectional analyses of a household panel survey. A separate cross-sectional file containing data for all responding households and their members (either deleting the households or imputing values for missing responses in the case of partial household nonresponse) can be created for each wave. Adjustments are then needed to compensate for the nonresponding and noncovered households and persons in each file.

Nonresponding households at wave t can be divided into total nonrespondents and wave nonrespondents. Total nonresponse occurs in a panel survey when a sampled element fails to provide data for any wave. Since it is common practice not to follow up sample households that fail to respond at the initial wave, these households and their members are generally the total nonrespondents. Compensation for total nonrespondents is relatively straightforward. The Wave 1 weights of the responding households at the initial wave can be adjusted using standard nonresponse adjustment methods and the adjusted weights can be used instead of the selection probabilities in developing the cross-sectional weights for later waves. Most nonresponse adjustment methods, such as weighting class adjustments (Kalton and Kasprzyk 1986) and adjustments based on response propensities (Little 1986), are based on the assumption that nonresponse is random within weighting classes or that the probabilities of responding within a class can estimated precisely. Under these conditions, the response mechanism can be treated as an additional stage of sampling. Thus, the selection probabilities, p_j, used to define the weights in equation (3.5) may be redefined as the product of the selection probabilities and the adjustment due to nonresponse. For example, if weighting class adjustments are used, the selection probability of original household h_j multiplied by the weighted response rate for the weighting class in which h_j falls is used instead of the original p_j. The previous results then follow for the weights adjusted for total nonresponse.

The same approach can also be extended to cover weighting adjustments for households responding at the initial wave that lead to no responding households at wave t. In this case, the responding households at the initial wave can be divided into weighting classes based on responses given at that wave, and the weights of households leading to one or more responding households at wave t

can be further adjusted to compensate for those leading to no responding households at wave t. The revised w_{ij}' can then be employed in equation (3.5) and subsequently.

Both the above nonresponse adjustments are applied in relation to the original households. A further type of household nonresponse cannot be handled in this way. This type of nonresponse involves the situation where an original household splits into two or more separate households at wave t, and where some but not all of those households respond at that wave. In this case the adjustment for the nonresponding households needs to be made in relation to the wave t households, H_i, rather than the original households, h_j. If the number of original households having members in each wave t nonresponding household of this type were known, the weights w_i for these households could be computed using the approach described above. Then weighting adjustments could be readily applied within weighting classes of the wave t households to compensate for the nonresponding households. In practice, however, the number of original households having members in a nonresponding household at wave t may often be unknown. One approach for handling this situation is to estimate this number by the average number for responding households at wave t that have similar characteristics to (*e.g.*, they are also splits from original households), and are in the same weighting class as, the nonresponding household. Using such estimated numbers where necessary, the weights w_i can be determined for all nonresponding households of the type being discussed. Standard weighting adjustments can then be applied to the responding households at wave t to compensate for these nonresponding households.

Incomplete coverage of the target population is another nonsampling problem that has been traditionally addressed in surveys by adjusting the sampling weights. For example, poststratification (see, for example, Holt and Smith 1979) and generalized raking procedures (Deville *et al.* 1993) are often used to adjust the weights so that they sum to counts from independent sources not subject to undercoverage. These adjustments may also reduce the sampling errors of the estimates, although bias reduction is often more critical.

The control totals used in most household surveys are counts of the number of persons in classes defined by characteristics such as age, sex and race. This method of reducing undercoverage bias may be fully sufficient when estimates of persons are the only types of statistics to be produced from the survey. However, further steps are needed to calculate household-level weights for producing statistics of household characteristics.

One approach to developing household-level weights when control totals are based on person-level counts is called the principal person method, as described by Alexander (1987). In this method, poststratification adjustments are applied at the person level. One household member is then identified as the principal person and the fully adjusted weight for that person is assigned to be the household weight. Since the person weights are already adjusted to the control totals, the household weight does incorporate some adjustments to reduce coverage bias. For cross-sectional estimation from a household panel survey, the principal person method can readily be used in conjunction with the equal household and person weighting schemes to produce household level weights.

A disadvantage of the principal person method is that estimates of the number of persons calculated using the principal person weight will generally differ from control totals. The estimates may also differ significantly depending on the criteria used to identify the principal person in the household. The method has also been criticized because the weights of the members of the same household differ, even though they were all selected at the same rate as the household. Estimation schemes proposed by Alexander (1987), Lemaître and Dufour (1987), and Zieschang (1990) address these objections by constraining the household weights so that they are consistent with the independent person-level totals while minimizing the distance between the original household weights and the adjusted weights. All three consider variants of a generalized least squares (GLS) algorithm to achieve this objective. Zieschang (1990) shows how GLS can be used to create weights that are consistent with the person controls and force all persons within a household to have the same weight.

The application of GLS methods when the household weights are computed using the equal household and person weighting schemes is relatively straightforward. However, empirical evaluation of the consequences of using these methods is needed. The GLS methods have the unattractive feature that they can result in negative weights. Furthermore, the increase in the variation in the weights arising from the constraints imposed may result in less precise estimates. This concern may be especially important when the variability in the household weights is increased due to their multiple routes for selection and the equal household or person weighting schemes are necessary.

7. SUMMARY AND CONCLUDING REMARKS

This paper has described weighting schemes for cross-sectional analysis of later waves of a household panel survey using data for all households for which and all individuals for whom data are collected. These weighting schemes can accommodate new entrants to the population who move in to live with members of the original population, but not other new entrants.

The usual inverse selection probability weighting scheme requires information on the household selection probabilities of all members of the households sampled at a later wave, as well as the joint selection probabilities of the original households that contribute members to the later wave households. The inverse selection probability weighting scheme can often not be applied because these probabilities are unknown. To deal with this problem, an alternative approach that requires information on only the selection probabilities of sampled original households is described.

This alternative approach produces a class of weighting schemes including the equal person (fair share) scheme used in SIPP and the equal household weighting scheme. All the schemes in this class produce weights that are in expectation equal to those produced by the usual inverse selection probability scheme. The variance in the weights around the inverse selection probability weights gives rise to an increase in the variance of the survey estimates. When the original households are selected with approximately equal probability, the equal household weighting scheme is near optimal for both household and individual level analyses to control this increase in variance.

The alternative class of weighting schemes produces unbiased estimates of population totals for any choice of constant α_{ij} that satisfies the condition $\sum_j \alpha_{ij} = 1$ and for any initial sample design. The equal household and equal person weighting schemes are, however, suboptimal for non-epsem initial samples. One of them may nevertheless be the appropriate scheme for such designs, because the optimal choice of the α_{ij} depends on the unknown initial selection probabilities, and hence cannot be determined. The equal household and equal person weighting schemes have different data requirements, in that the former requires knowledge of the number of Wave 1 households represented in the Wave t household whereas the latter does not. The fact that this information may not always be readily obtainable thus argues in favor of the equal person weighting scheme.

The cross-sectional individual weights for a particular wave can be used as the starting weights for a longitudinal analysis that begins at that wave. This procedure includes cohabitants present at that wave in the longitudinal analysis. However, if cohabitants are not followed when they cease to live with sampled persons, those who leave sample persons before the end of the period of the longitudinal analysis become nonrespondents. Before cohabitants are included in a longitudinal analysis, a check should therefore be made to ensure that their inclusion will not give rise to risks of serious nonresponse bias.

The class of weighting schemes described has a broader range of application than that indicated here. It can in fact be usefully applied in any situation where an inverse selection probability weighting scheme would be appropriate, but where not all the inclusion probabilities and joint inclusion probabilities are known. Consider, for instance, the modified version of the Mitofsky-Waksberg random digit dialing sampling procedure for telephone surveys described by Brick and Waksberg (1991). A sample of telephone numbers (primes) is selected at the first stage of this two-stage sample design. If a prime number is found to be a working residential number, that household is selected and a fixed number of additional telephone numbers in the same 100-bank is selected. The households found at these numbers are then all included in the sample. If a prime number is not a working number, the sampling process stops. With this procedure, the probability of a working residential number being selected depends on the number of working residential numbers in its 100-bank, and hence differs across 100-banks. This probability can be estimated from the sample of telephone numbers in the 100-bank. A complication arises, however, when a sampled household has two or more telephone numbers. In this case, the selection probability of the sampled telephone number can be estimated, but those of the nonsampled numbers cannot. Thus, the standard inverse selection probability weighting scheme cannot be used. However, the alternative weighting scheme described here can be employed.

ACKNOWLEDGEMENTS

We thank the referees for their helpful comments on the earlier version of the paper. The paper reports research undertaken for the U.S. Census Bureau. The views expressed are the authors'. They do not necessarily reflect the views of the Census Bureau.

REFERENCES

ALEXANDER, C.H. (1987). A class of methods for using person controls in household weighting. *Survey Methodology*, 13, 183-198.

BRICK, J.M., and WAKSBERG, J. (1991). Avoiding sequential sampling with random digit dialing. *Survey Methodology*, 17, 27-41.

BUCK, N., GERSHUNY, J., ROSE, D., and SCOTT, J. (Eds.) (1994). *Changing Households: The British Household Panel Survey 1990-1992.* Colchester, U.K.: ESRC Research Centre on Micro-social Change.

CITRO, C.F., and KALTON, G. (1993). *The Future of the Survey of Income and Program Participation.* Washington D.C.: National Academy Press.

DEVILLE, J.-C., SÄRNDAL, C.-E., and SAUTORY, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88, 1013-1020.

DUNCAN, G.J., and HILL, M.S. (1985). Conceptions of longitudinal households: Fertile or futile? *Journal of Economic and Social Measurement*, 13, 361-375.

ERNST, L.R. (1989). Weighting issues for longitudinal household and family estimates. In *Panel Surveys*, (Eds. D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh). New York: John Wiley, 139-159.

GAILLY, B., and LAVALLÉE, P. (1993). *Insérer des Nouveaux Membres dans un Panel Longitudinal de Ménages et D'Individus: Simulations*. Walferdange, Luxembourg: CEPS/Instead.

HILL, M.S. (1992). *The Panel Study of Income Dynamics: A User's Guide*. Newbury Park, CA: Sage Publications.

HILL, M.S. (1995). Personal Communication.

HOLT, D., and SMITH, T.M.F. (1979). Post stratification. *Journal of the Royal Statistical Society*, A, 142, 33-46.

HUANG, H. (1984). Obtaining cross-sectional estimates from a longitudinal survey: Experiences of the Income Survey Development Program. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 670-675.

JABINE, T.B., KING, K.E., and PETRONI, R.J. (1990). *Survey of Income and Program Participation: Quality Profile*. Washington D.C.: U.S. Bureau of the Census.

JUDKINS, D., HUBBLE, D., DORSCH, J., MCMILLEN, D., and ERNST, L. (1984). Weighting of persons for SIPP longitudinal tabulations. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 676-687.

KALTON, G. (1986). Handling wave nonrespone in panel surveys. *Journal of Official Statistics*, 2, 303-314.

KALTON, G., and KASPRZYK, D. (1986). The treatment of missing survey data. *Survey Methodology*, 12, 1-16.

KASPRZYK, D. (1988). *The Survey of Income and Program Participation: An Overview and Discussion of Research Issues*. SIPP Working Paper No. 8830. Washington D.C.: U.S. Bureau of the Census.

KASPRZYK, D., and MCMILLEN, D.B. (1987). SIPP: Characteristics of the 1984 Panel. *Proceedings of the Social Statistics Section, American Statistical Association*, 181-186.

KISH, L. (1965). *Survey Sampling*. New York: Wiley.

LAVALLÉE, P. (1995). Cross-sectional weighting of longitudinal surveys of individuals and households using the weight share method. *Survey Methodology*, 21, 25-32.

LAVALLÉE, P., and HUNTER, L. (1992). Weighting for the Survey of Labour and Income Dynamics. *Proceedings: Symposium 92, Design and Analysis of Longitudinal Surveys*, Statistics Canada, 65-75.

LAVALLÉE, P., MICHAUD, S., and WEBBER, M. (1993). The Survey of Labour and Income Dynamics, design issues for a new longitudinal survey in Canada. *Bulletin of the International Statistical Institute*, 49th Session, Contributed Papers, Book 2, 99-100.

LEMAÎTRE, G., and DUFOUR, J. (1987). An integrated method for weighting persons and families. *Survey Methodology*, 13, 199-207.

LEPKOWSKI, J. (1989). Treatment of wave nonresponse in panel surveys. In *Panel Surveys*, (Eds. D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh). New York: Wiley, 348-374.

LEPKOWSKI, J.M., MILLER, D.P., KALTON G., and SINGH, R. (1993). Imputation for wave nonresponse in the SIPP. *Proceedings: Symposium 92, Design and Analysis of Longitudinal Surveys*, Statistics Canada, 99-109.

LITTLE, R.J.A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review*, 54, 139-157.

LITTLE, R.J.A. (1989). Sampling weights in the PSID: Issues and comments. Panel Study of Income Dynamics Working Paper, Ann Arbor: University of Michigan.

NELSON, D., MCMILLEN, D., and KASPRZYK, D. (1985). *An Overview of the SIPP, Update* 1. SIPP Working Paper No. 8401. Washington D.C.: U.S. Bureau of the Census.

SINGH, R., HUGGINS, V., and KASPRZYK, D. (1990). *Handling Single Wave Nonresponse in Panel Surveys*. SIPP Working Paper No. 9009. Washington D.C.: U.S. Bureau of the Census.

ZIESCHANG, K.D. (1990). Sample weighting methods and estimation of totals in the Consumer Expenditure Survey. *Journal of the American Statistical Association*, 85, 986-1001.