ON THE WEIGHT SHARE METHOD FOR PANEL HOUSEHOLD SURVEYS

Takis Merkouris, Statistics Canada
Statistics Canada, R.H. Coats Bldg., 16-G, Tunney's Pasture, Ottawa, Ontario, K1A 0T6, Canada

Key Words: Cross-sectional estimation; cohabitants; interprovincial movers; longitudinal household; stratification; weight adjustment.

1. INTRODUCTION

A panel survey, though primarily conducted for longitudinal purposes, may also be used to produce cross-sectional estimates of population parameters at distinct time points (the survey waves). Cross-sectional weighting schemes that deal with dynamic aspects of a panel have been discussed in the literature. Kalton and Brick (1995) review such weighting schemes. Lavallée (1995) considers one of these weighting methods, termed *the weight share method*, in a more general context, and discusses its application to the Canadian Survey of Labour and Income Dynamics (SLID); for a description of SLID see Lavigne and Michaud (1998).

The weight share method is a cross-sectional weighting procedure that assigns a basic weight to every individual in a panel at any wave after the first. In particular, the weight share method assigns a positive weight to non-selected individuals who join households containing at least one individual selected for the original sample. Following Lavallée (1995), in this paper such households are termed longitudinal households, while the non-selected individuals living in longitudinal households are termed cohabitants. The cohabitants are distinguished into originally present cohabitants if they belong to the original (sampled) population, and originally absent cohabitants if they are new entrants to the population. Other problematic situations that can be handled by the weight share method involve households formed after the first wave by members of different originally selected households, as well as originally selected individuals who have subsequently moved to other longitudinal households.

This paper considers certain substantive aspects of the weight share method, as well as related practical issues. Specifically, since panel household surveys invariably employ stratification of the population at the time of selection of the sample, the alternative approach of applying the weight share procedure separately to each stratum is given particular attention. The case of practical interest involves a high level of stratification at which all other weighting and estimation procedures are carried out independently for each stratum. Such high level strata (superstrata) could be states or, as in the case of SLID, provinces. The characteristic feature of this alternative approach is that it treats as originally absent those individuals who at a subsequent survey wave reside in a stratum, say a province, other than the one in which they originally resided. The effect of applying the weight share method by province on statistical properties of derived estimators as well as the operational implications of this procedure are examined in contrast with the standard weight share method. The discussion is confined to single-panel household surveys, possibly supplemented with a "top-up" sample at some or all later survey waves. A top-up sample here means a new sample that covers the entire survey population at the time of sampling, but does not form a new panel. This sample is to be used only once, for cross-sectional purposes, and its size would normally be smaller than a panel's size.

A general formulation of the weight share method is presented first in Section 2. An account of implementation issues related to interprovincial movers is given in Section 3. The weight share method as applied by province, with elucidation of the unbiasedness of derived estimators at provincial and national levels, is described in Section 4. A formal comparison of the two approaches to the weight share adjustment is given in Section 5. The relative merits of the two weight share procedures in terms of coverage of the cross-sectional population, variance estimation and operational convenience are discussed in Section 6. Concluding remarks are made in Section 7.

2. A GENERAL FORMULATION OF THE WEIGHT SHARE METHOD

Let there be N individuals in the population at a survey wave (time t) after the selection of the panel, with N_i individuals in household $\mathcal{H}_i$ (i=1,...,H) and ΣN_i=N. Let M_i denote the number of individuals in household $\mathcal{H}_i$ that belong to the original (sampled) population U, with ΣM_i=M denoting the size of the remaining original population. One, but not both, of the numbers M_i and N_i-M_i may be zero for any particular household. For the individuals of the original population, the weights are defined as random variables that take the value of the inverse of the inclusion probability if the individuals are in the original sample, and the value of zero otherwise, whereas for individuals not in the original population the weights are defined to be equal to zero. Formally,

$$w_{ik} = \begin{cases} \frac{1}{\pi_{ik}} I(ik \in s), & \text{if } ik \in U \\ 0, & \text{if } ik \notin U , \end{cases}$$

where s is the panel sample, I is the usual sample membership indicator variable, and π_{ik} is the probability of inclusion in the sample for the k-th member of household $\mathcal{H}_i$. The weight share method defines a common weight for any individual in $\mathcal{H}_i$ as

$$w_i = \frac{1}{M_i} \sum_k^{M_i} w_{ik} = \begin{cases} \frac{1}{M_i} \sum_k^{M_i} \frac{1}{\pi_{ik}} I(ik \in s), & \text{if } M_i \neq 0 \\ 0, & \text{if } M_i = 0 , \end{cases} \qquad (1)$$

so that $E(w_i)=1$ for each i for which $M_i \neq 0$. If the inclusion probabilities are adjusted for nonresponse, the relationship $E(w_i)=1$ may hold only approximately.

For a survey characteristic y, the total for the population of individuals at time t can be expressed as

$$Y = \sum_i^H \sum_k^{N_i} y_{ik} = \sum_i^H \sum_k^{M_i} y_{ik} + \sum_i^H \sum_k^{N_i - M_i} y_{ik}$$

$$\doteq Y_o + Y_e ,$$

where y_{ik} is the value of y for the individual k in household $\mathcal{H}_i$. The two components Y_o and Y_e represent the total for the remaining of the original population and the total for the population of new entrants, respectively. Then, an estimator of Y is given by

$$\hat{Y} = \sum_i^H w_i \sum_k^{N_i} y_{ik} = \sum_i^H w_i \sum_k^{M_i} y_{ik} + \sum_i^H w_i \sum_k^{N_i - M_i} y_{ik} .$$

Note that households composed solely of new entrants (*i.e.*, with $M_i = 0$) are not represented in $\hat{Y}$. Then

$$E(\hat{Y}) = \sum_i^H \sum_k^{N_i} y_{ik} = \sum_i^H \sum_k^{M_i} y_{ik} + \sum_i^H \sum_k^{N_i - M_i} y_{ik}$$

$$\doteq Y_o + Y_{\bar{e}} ,$$

where $Y_{\bar{e}}$ denotes the total for the population of new entrants living in households that contain at least one member of the original population. Thus, unbiased estimators for both Y_o and $Y_{\bar{e}}$ are obtained, provided that the new entrants can be identified for the correct specification of M_i.

3. IMPLEMENTATION ISSUES

Dependence induced by movers between strata.

A complication arises in estimating variances of nation-level estimators that incorporate the weight share adjustment. The variance of a nation-level estimator cannot be obtained as the sum of the variances of the province-level estimators, as is readily done in cross-sectional surveys, because of the covariance terms induced by movers from one province to another. Nonzero covariance terms arise only among individuals that belong to the same stratum of the original province. Yet, the contribution of these covariances to the variance of the nation-level estimator may not be negligible.

The complication with calculation of the variances of nation-level estimators can be resolved by carrying out variance estimation at the nation level, treating the movers, for variance estimation purposes only, as still in their original province. Then estimates of variances at the province level can be obtained by treating provinces at time t as domains cutting across strata identified as the original provinces at the time of selection of the panel. For uncalibrated estimators this is a straightforward procedure for any of the resampling (replication) techniques usually employed for variance estimation in household surveys. For calibrated estimators based on (calibrated) person weights, the only additional requirement is the specification that the weights of the interprovincial movers are calibrated to the population control totals of their current province. This will slow down the variance estimation procedure considerably, as calibration has to be carried out simultaneously for all provinces for every replicate in the variance estimation procedure. For calibrated estimators based on integrated weights (a common calibrated weight within each household) the calibration algorithm becomes complicated, as it has to ensure that integrated weighting that satisfies the calibration constraints of each province is done properly in the original households of the interprovincial movers as well as in their new households, for every replicate in the variance estimation procedure. This may also slow down the variance estimation procedure even further.

It is important to point out here that the aforementioned problem of variance estimation arises at any level of stratum aggregation. The case of interest is variance estimation at the province level itself, carried out independently for the various provinces, as is customarily done in cross-sectional surveys. However, at the province level the variance estimation procedure works as prescribed above for variance estimation at nation level. Specifically, moves from stratum to stratum (or even from cluster to cluster) within a province are ignored for variance estimation purposes, and variances at stratum level are estimated, if needed, by treating the strata as domains. Difficulties relating to calibration, similar to those described above in the context of variance estimation at nation level, may be encountered also at the provincial level. Notwithstanding these complexities, at province level the weight share procedure can lead to a valid variance estimation procedure at affordable computational cost.

Differences in magnitude of weights between provinces. In cases where the magnitude of the weights of some interprovincial movers from a particular province of origin is much different from the magnitude of a typical weight in their new province, a common practice is an ad hoc adjustment of the weights of all movers from that province according to a known total accumulated number of movers from the same province of origin to the new province since the selection of the panel. This is not necessary for producing unbiased estimates of totals when the weight share procedure is used, even for the domain of movers. There are other reasons, however, for making such an adjustment, namely, nonresponse adjustment reasons, confidentiality concerns, and avoidance of erroneous calibrated weights (even negative ones in a multidimensional calibration), especially in small poststrata that contain such movers. Also, large differences between weights of movers and weights of original individuals in a province will most likely result in inflated variances of derived estimators, particularly for small domains containing movers. It is to be noted that such an adjustment of the weights of interprovincial movers can alleviate these problems, but cannot eliminate them. Moreover, the adjustment is not to be made at all if it would result in enlargement of the difference in magnitude between the weights of the movers and the weights of the original members of the movers' new province. It is to be noted further that accurate external information on the total accumulated number of interprovincial movers since the selection of the panel may not be available at each survey wave. Finally, given the large number of interprovincial move patterns (province of origin and province of destination), there is considerable operational complexity associated with this type of adjustment.

4. THE WEIGHT SHARE METHOD BY PROVINCE (PWS)

Consider the decomposition

$$Y=\sum_r^R Y_r=\sum_r^R\left(\sum_i^{H_r}\sum_k^{M_{ri}} y_{ik}+\sum_i^{H_r}\sum_k^{N_{ri}-M_{ri}} y_{ik}\right),$$

where Y_r is the total for the characteristic y in province P_r, H_r is the number of households in P_r at time t, M_{ri} is the number of the original individuals from P_r that are members of the household $\mathcal{H}_{ri}$ $(i=1,\ldots,H_r)$ at time t, and $N_{ri}-M_{ri}$ is the number of new entrants into P_r (including movers from other provinces) that are members of the household $\mathcal{H}_{ri}$ at time t. When the weight share procedure is applied by province all $N_{ri}-M_{ri}$ individuals are treated as originally absent in P_r (not having been selected there), and so their weights are set equal to zero, even for selected movers from other provinces. Thus, define a common weight for any individual in household $\mathcal{H}_{ri}$ in P_r as

$$w_{ri}=\frac{1}{M_{ri}}\sum_k^{M_{ri}} w_{ik}=\begin{cases}\dfrac{1}{M_{ri}}\displaystyle\sum_k^{M_{ri}}\frac{1}{\pi_{ik}}I(ik\in s_r), & \text{if } M_{ri}\neq 0\\ 0, & \text{if } M_{ri}=0\end{cases}$$

where s_r is the sample from the province P_r, drawn independently from other provinces with sampling design $p(s_r)$. Then, an estimator of Y_r is given by

$$\hat{Y}_r^{PWS}=\sum_i^{H_r} w_{ri}\sum_k^{M_{ri}} y_{ik}+\sum_i^{H_r} w_{ri}\sum_k^{N_{ri}-M_{ri}} y_{ik}.$$

Note that households in P_r with $M_{ri}=0$ are not represented in $\hat{Y}_r^{PWS}$. Such households may include individuals that are new entrants to the whole population, or movers from other provinces, or individuals from both of these categories. Now, $E_{p(s_r)}(w_{ri})=1$, for each i in P_r for which $M_{ri}\neq 0$. Then

$$E_{p(s_r)}(\hat{Y}_r^{PWS})=\sum_i^{H_r}\sum_k^{M_{ri}} y_{ik}+\sum_i^{H_r}\sum_k^{N_{ri}-M_{ri}} y_{ik}.$$

Thus, unbiased estimators are obtained for the remaining original population in P_r at time t, and for the population of new entrants (including movers from other provinces) into P_r living in households of P_r at time t that contain at least one member of the original population of P_r. Then, for $\hat{Y}^{PWS}=\sum_{r=1}^R \hat{Y}_r^{PWS}$,

$$E_{p(s)}(\hat{Y}^{PWS})=\sum_{r=1}^R E_{p(s_r)}(\hat{Y}_r^{PWS})=Y,$$

where $p(s)=\Pi_{r=1}^R p(s_r)$.

5. COMPARISONS OF THE TWO WEIGHT SHARE PROCEDURES

To compare the weight share procedure applied to the whole (nation) sample with the weight share procedure applied by province, rewrite the weight defined in (1) as

$$w_i = \frac{1}{M_i}\sum_k^{M_i} w_{ik} = \begin{cases} \dfrac{1}{\sum_{r=1}^{R} M_{ri}} \sum_{r=1}^{R}\sum_k^{M_{ri}} \dfrac{1}{\pi_{ik}} I(ik \in s_r), & \text{if } M_i \neq 0 \\ 0, & \text{if } M_i = 0 \end{cases}$$

for any individual in household $\mathcal{H}_{ri}$ in province P_r. Now $E_{p(s)}(w_i)=1$, for each i for which $M_i \neq 0$. Thus, unbiased estimators are obtained for the remaining original population in P_r at time t, for the entire population of interprovincial movers into P_r, and for the population of new entrants (e.g., immigrants) into P_r living in households of P_r at time t that contain at least one member of the entire original population (i.e., original members from P_r and movers into P_r).

A clear comparison of the two weight share procedures is afforded by rewriting estimated province-level totals for the two procedures in terms of the original design weights and household-level totals as follows.

(a) *Weight share for the national sample* (NWS)

The estimated total for P_r can take the form

$$\hat{Y}_r^{NWS} = \sum_i^{H_r}\sum_k^{M_{ri}} \frac{1}{\pi_{ik}} I(ik \in s_r)\frac{y_{ri}}{M_i} + \sum_i^{H_r}\sum_k^{M_i - M_{ri}} \frac{1}{\pi_{ik}} I(ik \in s \backslash s_r)\frac{y_{ri}}{M_i}$$

where

$$y_{ri} = \sum_k^{M_{ri}} y_{ik} + \sum_k^{N_{ri}-M_{ri}} y_{ik},$$

$M_i \neq 0$ for household $\mathcal{H}_{ri}$ in P_r, and $M_i - M_{ri}$ refers to the number of movers from the other provinces.

(b) *Weight share by province* (PWS).

The estimated total for P_r can take the form

$$\hat{Y}_r^{PWS} = \sum_i^{H_r}\sum_k^{M_{ri}} \frac{1}{\pi_{ik}} I(ik \in s_r)\frac{y_{ri}}{M_{ri}},$$

where y_{ri} is as in (a), and $M_{ri} \neq 0$ for household $\mathcal{H}_{ri}$ in P_r.

Notice the additional (statistically independent) term in the estimator based on the NWS, involving households that contain movers from the other provinces. Note also that terms associated with households for which $M_{ri}=0$ in P_r will be missing from the estimator based on the PWS procedure.

The two procedures differ in the construction of the household weights. Explicitly, the weight defined by the NWS procedure for members of household $\mathcal{H}_{ri}$ in province P_r can be expressed as $w_i = \sum_{r=1}^{R} c_r w_{ri}$, where $c_r = M_{ri}/M_i$, and w_{ri} is the household weight as defined for the PWS procedure in the previous section. Prior to the application of the PWS procedure, a zero weight is assigned to individuals who at time t reside in a province other than the one in which they originally resided. In effect, the PWS procedure treats these individuals as originally absent in their new province of residence at time t. In particular, movers (selected or non-selected in their original province) who are found in longitudinal households in their new province at time t are treated as originally absent cohabitants. On the other hand, the NWS procedure retains the original weights of the selected movers, and treats cohabitants coming from another province as originally present.

6. RELATIVE MERITS OF THE TWO PROCEDURES

It follows from the discussion in Section 5 that the implementation issues noted in Section 3 are resolved by using the PWS procedure. However, other issues are raised because of the treatment of the interprovincial movers by the PWS procedure. An account of the comparative merits of the two procedures is given below. For a more detailed discussion see Merkouris (1999 a).

Bias (coverage) considerations

In a single-panel survey, both procedures can estimate unbiasedly the same domain totals at province level, except that the PWS procedure cannot estimate the population of interprovincial movers who at time t live in households that contain no members of the original population of the movers' new province. In fact, the PWS procedure discards the selected movers of that type. In connection with the PWS procedure, the rest of the interprovincial movers are represented in the panel only through joining households that contain at least one selected individual from the original population, whereas in connection with the NWS procedure these interprovincial movers are sampled in their original province through the use of the frame at the time of the selection of the panel. Clearly, the hit rate for this type of interprovincial movers is lower with the PWS procedure.

The type of movers that is non-estimable by the PWS procedure constitutes a relatively very small domain within each province, which however may become sizeable over the lifetime of the panel for some provinces. Based on the first panel of SLID, selected in 1992, the accumulated number of these movers (estimated, using cross-sectional weights) over a three-year period represents 1.13% of the 1995 national population (10 provinces). By province, the percentage of these movers ranges from 0.28% of the 1995 Quebec population to 2.37% of the 1995 British Columbia population. Note that the maximum time period that may not be covered by a panel of SLID is three years, since a new panel is selected every three years. A calibration of the survey weights of the reduced sample (without these movers) to known population totals can lessen any bias effect of this type of noncoverage for characteristics correlated with the calibration variables. It is important to emphasize here that with a top-up sample at any survey wave the problematic domain in each province is covered, and thus it is estimable by the PWS procedure; for the combination of a panel and a top-up sample for cross-sectional estimation, see Merkouris (1999 b). It is to be noted that when a top-up sample is used, interprovincial movers (selected or non-selected in their original province) that are found in longitudinal households in their new province at time t are treated by the PWS procedure as originally present cohabitants.

It should be pointed out that interprovincial movers and their cohabitants in the new province are also discarded by the PWS procedure if their household does not contain selected members from the new province. This is because both types of household members have initial weights equal to zero. No bias is incurred in relation to the originally present cohabitants of the discarded movers in each province, but some bias may be associated with the originally absent cohabitants (*i.e.*, immigrants) of the discarded movers, since their population domain is not represented in the panel. This domain must be very small, as it is a rather rare event that new entrants into a province become cohabitants of interprovincial movers, and so the potential bias should be negligible.

In an empirical study of the differences between the NWS and PWS procedures, estimates were produced by each of the two procedures for several characteristics using data from the third wave of the first panel of SLID. In general, the differences were very small for most characteristics in most provinces. A few observed large differences correspond to population domains within which the estimated proportion of interprovincial movers into the particular province is much higher than the estimated overall proportion of interprovincial movers into the province; for example, the proportion of movers into New Brunswick that have income below the low income cut-off point (LICO) is more than three times larger than the overall proportion of movers into that province. The corresponding relative difference (with respect to the NWS procedure) for the number of individuals with income below the LICO is 6.1%. The potential for bias is, of course, larger for these domains. However, the few observed large relative differences do not necessarily indicate bias of the same magnitude. They may be explained to a large degree by sampling variability associated with interprovincial movers whose sampling weight is of much different magnitude from that of a typical weight in their new province; see relevant discussion in Section 3. At nation level the relative differences were very small. More information on this empirical study can be found in Merkouris (1999 a).

Variance considerations

In terms of efficiency of province-level estimators, an analytical assessment of the relative efficiencies of the two procedures is generally intractable for the part of the cross-sectional population that is estimable by both procedures. It is fair, though, to say that because of the very small number of households that contain movers from another province, the two procedures may not differ appreciably in terms of efficiency for this part of the population. Three years after the selection of the first panel of SLID, the accumulated number of selected movers in the panel who live with at least one member of the original population of their new province represents 0.49% of the total panel size. By province, the percentage of these movers ranges from 0.14% of the panel in Quebec to 1.20% of the panel in British Columbia. Non-selected movers that reside as cohabitants in longitudinal households are difficult to identify in SLID, but they should be very few. When a top-up sample is used, and the entire cross-sectional population is thus covered by both procedures, some loss of efficiency may be incurred by the PWS procedure due to discarding selected interprovincial movers (and their cohabitants) living in households with no member from the original population of the new province. This efficiency loss may become noticeable for some provinces over the lifetime of the panel, depending on the duration of the panel. In the first panel of SLID, the accumulated number of these movers over a three-year period represents 1.59% of the total panel size. By province, the percentage of these movers ranges from 0.41% of the panel in Quebec to 3.51% of the panel in British Columbia.

On the other hand, the NWS procedure may incur appreciable loss of efficiency if the differences between the weights of interprovincial movers of any type and the weights of units in the new province of the movers are large.

For an empirical comparison of the two weight share procedures in terms of efficiency of derived estimators, differences in CV's were also calculated in the study based on the SLID data. Because of the magnitude of the largest observed differences between estimates, and because the PWS procedure estimates a slightly smaller population, an assessment of the relative efficiencies based on CV's is more appropriate than the assessment based on variances. The difference in CV's was negligible over all provinces for most study characteristics. Note that the effect of the loss of the cohabitants of the discarded movers in the PWS procedure was also accounted for in the differences in CV's. It was interesting to notice a gain in efficiency in the same provinces and for almost all characteristics for which relative differences in estimates were large, despite the relatively high proportion of discarded units in the PWS procedure in these cases. This gain is effected partly because the PWS procedure avoids the inflationary effect on variances that is associated with interprovincial movers whose sampling weight is of much different magnitude from that of a typical weight in their new province.

Computation of variance estimates at both nation level and province level is feasible when the NWS procedure is used, but at a considerable operational complexity. In contrast, the PWS procedure retains the independence of the provincial samples, and thus nation-level variance estimates can then be readily obtained as sums of the province-level variance estimates.

Operational considerations

In terms of operational convenience, the PWS procedure is carried out in a straightforward manner. It only requires knowledge of whether a cohabitant came from another province in order to distinguish this cohabitant as originally absent. This distinction is not an issue when a top-up sample is used and the PWS procedure is applied after the combination of the panel and the top-up sample, for then all cohabitants are originally present; see Merkouris (1999 b). On the other hand, if the NWS procedure is to be applied, the weights of interprovincial movers may have to be adjusted (before the weight share) if the magnitude of these weights is much different from the magnitude of a typical weight in the movers' new province. This adjustment requires accurate external information on the accumulated number of interprovincial movers since the selection of the panel, which may not be readily available. There is also considerable operational complexity associated with such an adjustment. A great deal of additional operational complexity in the NWS procedure is associated with computation of variance estimates at national level.

7. CONCLUDING REMARKS

Certain complexities associated with interprovincial movers arise in the implementation of the weight share method. Applying the weight share method separately for each province eliminates those problems, but raises other issues because it involves discarding a certain type of interprovincial movers. The relative merits of the two approaches can be summarized in terms of operational convenience, bias and variance estimation as follows. The alternative weight share procedure is considerably more efficient operationally. It incurs very small bias, or no bias at all if a top-up sample is used. Its effect on efficiency is a negligible loss, or even gain for certain characteristics in some situations.

On a final point, when a supplementary sample of only new entrants (*e.g.*, immigrants) into the population is employed, the relative merits of the two weight share procedures are as in the case of a single-panel survey. For a panel survey scheme involving overlapping panels, the relative merits of the two approaches are as in the case of a single-panel supplemented with a top-up sample.

ACKNOWLEDGMENTS

The author wishes to thank Milorad Kovacevic and Douglas Yeo for their helpful comments that have led to the improvement of this paper.

REFERENCES

Kalton, G., and Brick, J.M. (1995). Weighting schemes for household panel surveys. *Survey Methodology*, 21, 33-44.

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

Lavigne, M., and Michaud, S. (1998). General aspects of the Survey of Labour and Income Dynamics. Working Paper SLID 98-05 E, Statistics Canada.

Merkouris, T. (1999 a). The weight share method for panel household surveys: Issues related to moves between strata. Methodology Branch Working Paper HSMD 99-003E, Statistics Canada.

Merkouris, T. (1999 b). Cross-sectional estimation in multiple-panel household surveys. Methodology Branch Working Paper HSMD 99-004E, Statistics Canada.