Methodology Branch

Direction de la méthodologie

Household Survey Methods Division

Division des méthodes d'enquêtes auprès des ménages

Statistics Canada Statistique Canada

Canada

WORKING PAPER
METHODOLOGY BRANCH

THE WEIGHT SHARE METHOD FOR PANEL HOUSEHOLD SURVEYS: ISSUES RELATED TO MOVES BETWEEN STRATA

HSMD - 99 - 003E

Takis Merkouris

Household Survey Methods Division
Statistics Canada

October 1999

The work presented in this paper is the responsibility of the author and does not necessarily represent the views or policies of Statistics Canada

The Weight Share Method for Panel Household Surveys: Issues Related to Moves Between Strata

TAKIS MERKOURIS[1]

ABSTRACT

When a panel survey is used for cross-sectional purposes, a weighting adjustment procedure, known as *the weight share method*, is employed to deal with dynamic aspects of the panel, such as movers and cohabitants (Kalton and Brick 1995, Lavallée 1995). In this paper, the merits of applying the weight share method separately to each sample stratum at the highest level of stratification are investigated. Such high level strata (superstrata) could be, for example, states or provinces. This approach is characterized by a redefinition of the concepts of mover and cohabitant that is relative to the superstrata rather than to the entire population. The impact of applying the weight share method separately by superstratum on statistical properties of cross-sectional estimators as well as the operational ramifications of this procedure are examined in contrast with the standard weight share method. The comparative merits of the two approaches are assessed under different panel survey schemes.

KEY WORDS: Cross-sectional estimation; cohabitants; interprovincial movers; longitudinal household; stratification; weight adjustment.

Takis Merkouris, Survey and Analysis Methods Development Section, Household Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6.

Méthode du partage des poids appliquée aux enquêtes-ménages par panel: Questions reliées à la inter-strates migration

TAKIS MERKOURIS[2]

RÉSUMÉ

Lorsqu'on utilise une enquête par panel pour des observations transversales, on a recours à une procédure d'ajustement des poids, appelée *la méthode du partage des poids*, pour traiter les éléments dynamiques du panel, notamment les personnes ayant déménagé et les cohabitants (Kalton and Brick 1995, Lavallée 1995). La présente étude porte sur les mérites d'appliquer séparément la méthode du partage des poids à chaque strate échantillonnée au plus haut niveau de stratification. Ces strates supérieures pourraient être, par exemple, des États ou des provinces. Cette méthode est caractérisée par une redéfinition des concepts de personnes ayant déménagé et de cohabitants qui est relative aux strates supérieures plutôt qu'à l'ensemble de la population. L'étude examine l'incidence de l'application de la méthode du partage des poids à chaque strate supérieure sur les propriétés statistiques des estimateurs transversaux ainsi que les ramifications opérationnelles de cette procédure, comparativement à la méthode standard de partage des poids. Les mérites des deux méthodes sont ensuite évalués suivant différents scénarios d'enquête par panel.

MOTS CLÉS: Estimation transversale; cohabitants; migrants interprovinciaux; ménage longitudinal; stratification; ajustement des poids.

Takis Merkouris, Section du dévelopment des méthodes d' enquêtes et d'analyse, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada, Ottawa, Ontario, K1A 0T6.

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). The process of obtaining cross-sectional estimates at any wave of a panel household survey after the first wave presents difficulties arising from the dynamic character of the panel. Weighting schemes which 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 in detail 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 deserves 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. In particular, individuals (selected or non-selected in their original province) that are found in longitudinal households in their new province at a subsequent survey wave are treated as originally absent cohabitants. 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. 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 3. In Section 4, the standard weight share method is reworked along the lines of Section 3 for comparative purposes. The two weight share procedures are compared in Section 5 in terms of coverage of the cross-sectional population, variance estimation and operational convenience. A summary of the relative merits of the two methods is given in Section 6.

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 $\sum N_i = N$. Let M_i denote the number of individuals in household $\mathcal{H}_i$

that belong to the original (sampled) population U, with $\Sigma M_i = M$ 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,

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

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

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

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

$$\begin{aligned} Y &= \sum_i^H \sum_k^{N_i} y_{ik} = \sum_i^H \sum_k^{M_i} y_{ik} + \sum_i^H \sum_k^{N_i - M_i} y_{ik} \\ &\doteq Y_o + Y_e , \end{aligned}$$

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

$$\begin{aligned} E(\hat{Y}) &= \sum_i^H \sum_k^{N_i} y_{ik} = \sum_i^H \sum_k^{M_i} y_{ik} + \sum_i^H \sum_k^{N_i - M_i} y_{ik} \\ &\doteq Y_o + Y_{\bar{e}} . \end{aligned}$$

where $Y_{\bar{e}}$ is 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.
An alternative expression for $\hat{Y}$ in terms of the household totals and the original design weights is given by

$$\hat{Y}=\sum_i^H \sum_k^{M_i} w_{ik}\frac{y_i}{M_i}=\sum_i^H \sum_k^{M_i} w_{ik}\frac{y_{oi}}{M_i}+\sum_i^H \sum_k^{M_i} w_{ik}\frac{y_{\bar{e}i}}{M_i},$$

where

$$y_i=\sum_k^{N_i} y_{ik}=\sum_k^{M_i} y_{ik}+\sum_k^{N_i-M_i} y_{ik}$$
$$=y_{oi}+y_{\bar{e}i}.$$

Clearly, the estimator $\hat{Y}$ is given as a sum of two Horwitz-Thompson estimators, for the household characteristics y_{oi}/M_i and $y_{\bar{e}i}/M_i$, related to the original population and the population of new entrants.

3. THE WEIGHT SHARE METHOD BY PROVINCE (PWS)

For simplicity, assume there are only two provinces, denoted P1 and P2, with totals Y_1 and Y_2, respectively, for the characteristic y. Consider then the decomposition

$$Y=Y_1+Y_2=\left(\sum_i^{H_1}\sum_k^{M_{1i}} y_{ik}+\sum_i^{H_1}\sum_k^{N_{1i}-M_{1i}} y_{ik}\right)+\left(\sum_i^{H_2}\sum_k^{M_{2i}} y_{ik}+\sum_i^{H_2}\sum_k^{N_{2i}-M_{2i}} y_{ik}\right),$$

where H_1 is the number of households in P1 at time t, M_{1i} is the number of the original individuals from P1 that are members of the household $\mathcal{H}_{1i}$ at time t, and $N_{1i}-M_{1i}$ is the number of new entrants into P1 (including movers from P2) that are members of the household $\mathcal{H}_{1i}$ at time t. When the weight share procedure is applied by province all $N_{1i}-M_{1i}$ individuals are treated as originally absent in P1 (not having been selected there), and so their weights are set equal to zero, even for selected movers from P2. The corresponding notations for P2 are similar. Next, define a common weight for any individual in household $\mathcal{H}_{1i}$ in P1 as

$$w_{1i}=\frac{1}{M_{1i}}\sum_k^{M_{1i}} w_{ik}=\begin{cases}\dfrac{1}{M_{1i}}\displaystyle\sum_k^{M_{1i}}\frac{1}{\pi_{ik}}I(ik\in s_1), & \text{if } M_{1i}\neq 0\\ 0, & \text{if } M_{1i}=0\end{cases}$$

and a common weight for any individual in household $\mathcal{H}_{2i}$ in P2 as

$$w_{2i}=\frac{1}{M_{2i}}\sum_{k}^{M_{2i}} w_{ik}=\begin{cases}\frac{1}{M_{2i}}\sum_{k}^{M_{2i}}\frac{1}{\pi_{ik}}I(ik\in s_2), & if\ M_{2i}\neq 0\\ 0, & if\ M_{2i}=0\ ,\end{cases}$$

where s_1 and s_2 are the samples from the two provinces, independently drawn with sampling designs $p(s_1)$ and $p(s_2)$. Then, estimators for Y_1 and Y_2 are given, respectively, by

$$\hat{Y}_1^{PWS}=\sum_{i}^{H_1} w_{1i}\sum_{k}^{M_{1i}} y_{ik}+\sum_{i}^{H_1} w_{1i}\sum_{k}^{N_{1i}-M_{1i}} y_{ik},$$

and

$$\hat{Y}_2^{PWS}=\sum_{i}^{H_2} w_{2i}\sum_{k}^{M_{2i}} y_{ik}+\sum_{i}^{H_2} w_{2i}\sum_{k}^{N_{2i}-M_{2i}} y_{ik}.$$

Note that households in P1 with $M_{1i}=0$ are not represented in $\hat{Y}_1$. Such households may include individuals that are new entrants to the whole population, or movers from P2, or individuals from both of these categories. Similarly for P2. Now, $E_{p(s_1)}(w_{1i})=1$, for each i in P1 for which $M_{1i}\neq 0$, and $E_{p(s_2)}(w_{2i})=1$, for each i in P2 for which $M_{2i}\neq 0$. Then

$$E_{p(s_1)}(\hat{Y}_1^{PWS})=\sum_{i}^{H_1}\sum_{k}^{M_{1i}} y_{ik}+\sum_{i}^{H_1}\sum_{k}^{N_{1i}-M_{1i}} y_{ik},$$

and

$$E_{p(s_2)}(\hat{Y}_2^{PWS})=\sum_{i}^{H_2}\sum_{k}^{M_{2i}} y_{ik}+\sum_{i}^{H_2}\sum_{k}^{N_{2i}-M_{2i}} y_{ik}.$$

Thus, unbiased estimators are obtained for the remaining original population in P1 at time t, and for the population of new entrants (including movers from P2) into P1 living in households of P1 at time t that contain at least one member of the original population of P1. Similarly for P2. Then

$$E_{p(s)}(\hat{Y}^{PWS})=\left(\sum_{i}^{H_1}\sum_{k}^{M_{1i}} y_{ik}+\sum_{i}^{H_1}\sum_{k}^{N_{1i}-M_{1i}} y_{ik}\right)+\left(\sum_{i}^{H_2}\sum_{k}^{M_{2i}} y_{ik}+\sum_{i}^{H_2}\sum_{k}^{N_{2i}-M_{2i}} y_{ik}\right), \quad (2)$$

where $p(s)=p(s_1)p(s_2)$.

4. THE WEIGHT SHARE METHOD FOR THE NATIONAL SAMPLE (NWS)

Consider again the setting involving the two provinces, P1 and P2, but now with the decomposition

$$Y = Y_1 + Y_2 = \left[\sum_i^{H_1}\sum_k^{M_i} y_{ik} + \sum_i^{H_1}\sum_k^{N_{1i}-M_i} y_{ik}\right] + \left[\sum_i^{H_2}\sum_k^{M_i} y_{ik} + \sum_i^{H_2}\sum_k^{N_{2i}-M_i} y_{ik}\right]$$

$$= \left[\sum_i^{H_1}\left(\sum_k^{M_{1i}} y_{ik} + \sum_k^{M_{2i}} y_{ik}\right) + \sum_i^{H_1}\sum_k^{N_{1i}-M_i} y_{ik}\right] + \left[\sum_i^{H_2}\left(\sum_k^{M_{2i}} y_{ik} + \sum_k^{M_{1i}} y_{ik}\right) + \sum_i^{H_2}\sum_k^{N_{2i}-M_i} y_{ik}\right],$$

where M_{1i} and M_{2i} are as before (with $M_{1i}+M_{2i}=M_i$).

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

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

$$= \begin{cases} \dfrac{1}{M_{1i}+M_{2i}}\left[\displaystyle\sum_k^{M_{1i}} \frac{1}{\pi_{ik}} I(ik \in s_1) + \sum_k^{M_{2i}} \frac{1}{\pi_{ik}} I(ik \in s_2)\right], & \text{if } M_i \neq 0 \\ 0, & \text{if } M_i = 0 \end{cases}$$

for any individual in household $\mathcal{H}_{1i}$ in province P1, or in household $\mathcal{H}_{2i}$ in province P2. In the above expression, M_{1i} or M_{2i}, depending on the province of the household, refers to the number of movers from the other province. Now $E_{p(s)}(w_i)=1$, for each i for which $M_i \neq 0$. Note that $w_i = cw_{1i} + (1-c)w_{2i}$, where $c=M_{1i}/M_i$, and w_{1i}, w_{2i} are the weights (for household members from the original population of P1 and P2, respectively) defined for the PWS procedure in the previous section. Estimators for Y_1 and Y_2 are now given by

$$\hat{Y}_1^{NWS} = \sum_i^{H_1} w_i\left(\sum_k^{M_{1i}} y_{ik} + \sum_k^{M_{2i}} y_{ik}\right) + \sum_i^{H_1} w_i \sum_k^{N_{1i}-M_i} y_{ik},$$

and

$$\hat{Y}_2^{NWS} = \sum_i^{H_2} w_i \left(\sum_k^{M_{2i}} y_{ik} + \sum_k^{M_{1i}} y_{ik} \right) + \sum_i^{H_2} w_i \sum_k^{N_{2i}-M_i} y_{ik},$$

respectively. It follows that

$$E_{p(s)}(\hat{Y}^{NWS}) = E_{p(s)}(\hat{Y}_1^{NWS}) + E_{p(s)}(\hat{Y}_2^{NWS})$$

$$= \left[\sum_i^{H_1} \left(\sum_k^{M_{1i}} y_{ik} + \sum_k^{M_{2i}} y_{ik} \right) + \sum_i^{H_1} \sum_k^{N_{1i}-M_i} y_{ik} \right] + \left[\sum_i^{H_2} \left(\sum_k^{M_{2i}} y_{ik} + \sum_k^{M_{1i}} y_{ik} \right) + \sum_i^{H_2} \sum_k^{N_{2i}-M_i} y_{ik} \right] \quad (3)$$

$$= \left[\sum_i^{H_1} \sum_k^{M_{1i}} y_{ik} + \sum_i^{H_1} \sum_k^{N_{1i}-M_{1i}} y_{ik} \right] + \left[\sum_i^{H_2} \sum_k^{M_{2i}} y_{ik} + \sum_i^{H_2} \sum_k^{N_{2i}-M_{2i}} y_{ik} \right].$$

Thus, unbiased estimators are obtained for the remaining original population in P1 at time t , for the entire population of interprovincial movers into P1, and for the population of new entrants (e.g., immigrants) into P1 living in households of P1 at time t that contain at least one member of the entire original population (i.e., original members from P1 and movers into P1). Similarly for P2.

5. COMPARISONS OF THE TWO WEIGHT SHARE PROCEDURES

The two procedures differ in the construction of the household weights. Specifically, 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. Obviously, the two procedures would be equivalent in the trivial case in which all individuals (selected or cohabitants) were in their original province at time t .

Bias considerations

In a single-panel survey, both procedures can estimate unbiasedly the same domain totals (compare expressions (2) and (3)) 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 though may become sizeable over the life time 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 cannot 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. In such a case, unbiased estimation of this domain requires that its members in the top-up sample be identified (as new entrants into their new province), so that the dual-frame type of weight adjustment required for combining the panel and the top-up sample does not apply to them; for the combination of a panel and a top-up sample for cross-sectional estimation, see Merkouris (1999). Identification of the interprovincial movers in the top-up sample may not be possible under the operational procedures of a panel household survey. In that case the non-identifiable domain of movers would be underestimated. 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. On the other hand, with a top-up sample the NWS procedure produces unbiased estimators for this domain of movers always, regardless of the identifiability of such movers in the top-up sample. This is because the interprovincial movers are not treated by the NWS procedure as new entrants.

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 are cohabitants of interprovincial movers, and so the potential bias should be negligible. Furthermore, the benefits derived from the calibration of the sampling weights or from the use of a top-up sample, regarding potential bias due to discarding of interprovincial movers, also apply to the cohabitants of these movers.

Table 1 below shows the percent relative differences in estimates produced by the NWS and PWS procedures using data from the third wave of the first panel of SLID. The relative differences are with respect to the NWS procedure. The reported variables are those for which the largest differences were observed; they are, the number of individuals with the marital status of common law, the number of individuals with income below the low income cut-off point (LICO), the total (provincial and national) person income, the total income for families of size one ("unattached" persons), and the corresponding average incomes. For each variable the minimum and maximum relative differences by province are shown, as well as the overall relative difference at national level. The relative differences in the other provinces are closer to the minimum than to the maximum shown in Table 1. The 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 LICO is more than three times larger than the overall proportion of movers into that province. 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 later in this section. Note that at Canada level the relative differences are very small.

Table 1. *Relative differences in estimates (%)*

Variable	Min.	By Prov.	Max.	By Prov.	Canada
Common law	0.05	MAN	12.68	BC	1.76
LICO	0.16	PEI	6.12	NB	1.35
Income	0.01	SK	1.41	AB	-0.03
Income Fs1	-0.06	MAN	8.44	AB	1.50
Av. Income	0.02	QUE	1.35	AB	-0.04
Av. Inc. Fs1	0.03	MAN	6.50	AB	0.25

Variance considerations

For comparison purposes, let us rewrite estimated household-level and province-level totals for the two weight share procedures in terms of the original design weights as follows.

(a) NWS

For household $\mathcal{H}_{1i}$ in P1 for which $M_i \neq 0$ the estimated total can be written as

$$\sum_{k}^{M_i} \frac{1}{\pi_{ik}} I(ik \in s) \frac{y_{1i}}{M_i} = \sum_{k}^{M_{1i}} \frac{1}{\pi_{ik}} I(ik \in s_1) \frac{y_{1i}}{M_i} + \sum_{k}^{M_{2i}} \frac{1}{\pi_{ik}} I(ik \in s_2) \frac{y_{1i}}{M_i},$$

where

$$y_{1i} = \sum_{k}^{M_{1i}} y_{ik} + \sum_{k}^{N_{1i}-M_{1i}} y_{ik},$$

so that the estimated total for P1 can take the form

$$\hat{Y}_1^{NWS} = \sum_{i}^{H_1} \sum_{k}^{M_{1i}} \frac{1}{\pi_{ik}} I(ik \in s_1) \frac{y_{1i}}{M_i} + \sum_{i}^{H_1} \sum_{k}^{M_{2i}} \frac{1}{\pi_{ik}} I(ik \in s_2) \frac{y_{1i}}{M_i}. \quad (4)$$

Similarly, for household $\mathcal{H}_{2i}$ in P2 for which $M_i \neq 0$ the estimated total can be written as

$$\sum_{k}^{M_i} \frac{1}{\pi_{ik}} I(ik \in s) \frac{y_{2i}}{M_i} = \sum_{k}^{M_{2i}} \frac{1}{\pi_{ik}} I(ik \in s_2) \frac{y_{2i}}{M_i} + \sum_{k}^{M_{1i}} \frac{1}{\pi_{ik}} I(ik \in s_1) \frac{y_{2i}}{M_i},$$

where

$$y_{2i} = \sum_{k}^{M_{2i}} y_{ik} + \sum_{k}^{N_{2i}-M_{2i}} y_{ik},$$

so that

$$\hat{Y}_2^{NWS} = \sum_{i}^{H_2} \sum_{k}^{M_{2i}} \frac{1}{\pi_{ik}} I(ik \in s_2) \frac{y_{2i}}{M_i} + \sum_{i}^{H_2} \sum_{k}^{M_{1i}} \frac{1}{\pi_{ik}} I(ik \in s_1) \frac{y_{2i}}{M_i}. \qquad (5)$$

(b) PWS

For household $\mathcal{H}_{1i}$ in P1 for which $M_{1i} \neq 0$ the estimated total can be written as

$$\sum_{k}^{M_{1i}} \frac{1}{\pi_{ik}} I(ik \in s_1) \frac{y_{1i}}{M_{1i}},$$

where y_{1i} is as in (a). Then

$$\hat{Y}_1^{PWS} = \sum_{i}^{H_1} \sum_{k}^{M_{1i}} \frac{1}{\pi_{ik}} I(ik \in s_1) \frac{y_{1i}}{M_{1i}}.$$

Similarly, for household $\mathcal{H}_{2i}$ in P2 for wich $M_{2i} \neq 0$ the estimated total can be written as

$$\sum_{k}^{M_{2i}} \frac{1}{\pi_{ik}} I(ik \in s_2) \frac{y_{2i}}{M_{2i}},$$

where y_{2i} is as in (a), so that

$$\hat{Y}_2^{PWS} = \sum_{i}^{H_2} \sum_{k}^{M_{2i}} \frac{1}{\pi_{ik}} I(ik \in s_2) \frac{y_{2i}}{M_{2i}}.$$

The above expressions for the estimators constructed by the two weight share procedures make the differences of these procedures explicit. The prime difference is the additional (statistically independent) term in the estimator based on the procedure (a), involving households that contain movers from the other province. On the other hand, within households of a province the terms associated with the province's sample are larger for procedure (b) than the corresponding ones for procedure (a) by the factor M_i / M_{1i} for household $\mathcal{H}_{1i}$ in P1, and by the factor M_i / M_{2i} for household $\mathcal{H}_{2i}$ in P2, due to the treatment in procedure (b) of all new entrants to a province as originally absent. Note that terms associated with households for which M_{1i}=0 in P1 and M_{2i}=0 in P2 will be missing from the estimator based on procedure (b).

In a given panel survey, with samples s_1 and s_2 independently drawn in P1 and P2, respectively, a large difference in the magnitude of the probabilities of inclusion in s_1 and s_2 may result in the totals of households that contain movers from another province being larger or smaller for one of the two procedures. This, in turn, may result in the variances of these household totals being larger or smaller for one of the two procedures. But with covariance terms across households within each of the samples s_1 and s_2 it is not in general possible to assess analytically which of the two procedures produces more efficient estimators for Y_1 and Y_2. Nevertheless, the number of households in the panel containing movers from another province is likely to be very small, and thus the two procedures may be only slightly different in efficiency, with respect to the part of the cross-sectional population that is estimable by both procedures. 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, relative to the NWS procedure, due to discarding households composed solely of movers selected in another province. 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.

Both weight share procedures introduce variability into population estimates because the selection probabilities of new members that have joined longitudinal households are generally unknown; see Kalton and Brick (1995), and Czajka (1994) for relevant discussion. The amount of this variability increases with the variability in the selection probabilities of the members of such households. If, however, a new member in a longitudinal household is an originally absent cohabitant (and as such has zero weight) the addition of this new member has no impact on the household weight; the weight is identical to what it would have been if this individual had not joined the household. In a single-panel survey, since the interprovincial movers that are found in longitudinal households in their new province are treated as originally absent cohabitants by the PWS procedure, the variability introduced into population estimates by the PWS procedure is less than the variability introduced by the NWS procedure, especially when there is a large difference in the magnitude of the weights of the interprovincial movers and the weights of the original members in the movers' new province. It is to be noted that this particular comparison is only in relation to the additional component of variability (conditional on the set of longitudinal households in the panel at time t, and averaged over all such possible sets) resulting from the use of either weight share procedure.

For an empirical comparison of the two weight share procedures in terms of efficiency of derived estimators, Table 2 below shows the difference in CV's (which are in percent) for the same survey characteristics for which the relative differences in the estimates produced by the two procedures are large. Because of the magnitude of the differences between these estimates, and since 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. For each of the variables, the numbers in Table 2 represent the maximum loss or maximum gain in terms of CV, by province, associated with the use of the alternative PWS procedure.

Table 2. *Differences in CV's*

Variable	**Max. Loss**		**Max. Gain**	
Common law	0.88	AB	1.06	BC
LICO	0.61	NB		
Income	0.07	BC	0.44	AB
Income Fs1	0.53	PEI	3.00	AB
Av. Income	0.09	PEI	0.45	AB
Av. Inc. Fs1	0.56	PEI	3.98	AB

The maximum loss is very small in all reported cases, while the difference in CV's is negligible over all provinces for many other characteristics not reported here. Note that the effect of the loss of the cohabitants of the discarded movers in the PWS procedure is also accounted for in the differences in CV's. It is interesting to notice the gain in efficiency in the same provinces and for almost all characteristics for which relative differences in estimates were large in Table 1, despite the relatively high proportion of discarded units in the PWS procedure in these cases. This should be 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.

The expressions (4) and (5) point to a complication in estimating variances of national-level estimators that incorporate the weight adjustment based on the NWS procedure. The variance of a national-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. As evident from expressions (4) and (5), there is nonzero covariance between the first term of $\hat{Y}_1^{NWS}$ and the second term of $\hat{Y}_2^{NWS}$, as well as between the first term of $\hat{Y}_2^{NWS}$ and the second term of $\hat{Y}_1^{NWS}$. Nonzero covariance terms arise only among individuals that belong to the same stratum of the original province. Nevertheless, the second term in $\hat{Y}_1^{NWS}$ and $\hat{Y}_2^{NWS}$ includes all types of movers, and thus the contribution of these covariances to the variance of the national-level estimator $\hat{Y}_1^{NWS}+\hat{Y}_2^{NWS}$ may not be negligible. The effect of ignoring the covariances among provinces in calculating national-level variances was evaluated using data from the second wave of the National Population Health Survey (NPHS). Table 3 below shows the percent overestimation (negative numbers) or underestimation of the variance of totals for the indicated characteristics. While the effect on the two health characteristics, shown in the column margins of the table, is not substantial, the effect on the auxiliary characteristics, shown in the row margins, is considerable. The effect on the health characteristics is more pronounced when they are crossed with the auxiliary ones. As the number of interprovincial movers usually increases with the life time of the panel, such effects of ignoring the covariances between provinces are also likely to increase.

Table 3. *Overestimation or underestimation of national-level variances (%).*

Variable	Restricted Activity	Smoking	
Household size one	1.0	-5.1	3
Common Law	-6.1	-6.4	-11
Separated/Divorced	-3.6	5	2.7
Single	0.6	-3.7	-3.1
Student	3.8	4.6	6.8
	-1.8	2.4	

The complication with calculation of the variances of national-level estimators can be resolved by carrying out variance estimation at national level, treating the movers as still in their original province, only for variance estimation purposes. Then estimates of variances at 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. The complication arises because the movers are to be resampled in their original province, but in a manner whereby they are not to be counted as members of their original household. At the same time, these movers must be integrated for calibration purposes with the other members (if any) of their new household. This may also slow down the variance estimation procedure even further. In the case of NPHS, although a very simple calibration scheme (poststratification in 16 age/sex groups) without integration of weights is used, the bootstrap variance estimation procedure is computationally intensive. It is interesting to note that for the replicate that does not include the original primary sampling unit (PSU) associated with one or more interprovincial movers, these movers are to be treated in their new province as originally present cohabitants.

It is important to point out here that the aforementioned problem of variance estimation arises at any level of strata 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. Then the problem arises also in connection with the PWS procedure. However, at the province level the variance estimation procedure works as prescribed above for variance estimation at national 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 national level, may be encountered also at the provincial level. An additional operational complexity is introduced by the NWS procedure, as the variability associated with the interprovincial movers must be incorporated into the variance estimation procedure to derive valid provincial variance estimators. These complexities notwithstanding, at province level both weight share procedures can lead to valid variance estimation procedures at affordable computational cost.

Other Considerations

In connection with the NWS procedure, 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, the practice in SLID 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. In view of the findings in Section 2 and Section 3 this is not necessary for producing unbiased estimates of totals when the NWS 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.

Clearly, the problems arising from large differences in magnitude between the weights of interprovincial movers and the weights of original members in a province are avoided by using the PWS procedure.

6. SUMMARY OF THE RELATIVE MERITS OF THE TWO WEIGHT SHARE PROCEDURES

The relative merits of the two weight share procedures can be summarized in terms of bias, variance estimation and operational convenience as follows.

In terms of bias (more precisely, coverage), in a single-panel survey the PWS procedure cannot produce estimates for 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. Characteristics of this small population domain in each province would be unbiasedly estimated by the PWS procedure if a top-up sample were used at any survey wave, provided that members of this population domain could be identified in the top-up sample.

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. 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 life time of the panel, depending on the duration of the panel. 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.

Variance estimation at the province level is somewhat more complicated when the NWS procedure is used, since the variability associated with the interprovincial movers must be incorporated into the variance estimation procedure to derive valid province-level variance estimators. Furthermore, variance estimators at national level cannot be obtained as sums of the province-level variance estimators when the NWS procedure is used, because of the covariances induced by movers from one province to another. Computation of variance estimates at both national level and province level is feasible, but comes at a considerable operational complexity. In contrast, the PWS procedure retains the independence of the provincial samples, and thus national-level variance estimates can then be readily obtained as sums of the province-level variance estimates.

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). 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.

Finally, it is to be noted that 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 Harold Mantel and Michel Latouche for their helpful comments that have led to the improvement of this paper. Special thanks also to Harold Mantel for providing the numbers of Table 3, to Christian Nadeau and Eric Gagnon for providing the data and the estimation programs of SLID, and to Carleton University student JingHong Hu for generating the numbers of Table 1 and Table 2.

REFERENCES

Czajka, J.L. (1994). Income stratification in panel surveys: Issues in design and estimation. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 791-796.

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