Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Treatments for link nonresponse in indirect sampling

by Xiaojian Xu and Pierre Lavallée

December 2009

Statistics Canada Statistique Canada

Canada

Survey Methodology, December 2009
Vol. 35, No. 2, pp. 153-164
Statistics Canada, Catalogue No. 12-001-X

Treatments for link nonresponse in indirect sampling

Xiaojian Xu and Pierre Lavallée [1]

Abstract

We examine overcoming the overestimation in using generalized weight share method (GWSM) caused by link nonresponse in indirect sampling. A few adjustment methods incorporating link nonresponse in using GWSM have been constructed for situations both with and without the availability of auxiliary variables. A simulation study on a longitudinal survey is presented using some of the adjustment methods we recommend. The simulation results show that these adjusted GWSMs perform well in reducing both estimation bias and variance. The advancement in bias reduction is significant.

Key Words: Weight share method; Nonresponse; Indirect sampling; Longitudinal survey.

1. Introduction

Indirect sampling refers to selecting samples from the population which is not, but it is related to, the target population of interest. Such a sampling scheme is often carried out when we do not have sampling frames for the target population, but have sampling frames for another population which is related to it. We call the latter sampling population. For an example in Lavallée (2007), we consider the situation where the estimate is concerned with young children belonging to families, but we only have a list of parents' names as our sampling frame. Consequently, we must first select a sample of parents before we can select the sample of children. In this typical indirect sampling situation. The sampling population is that of parents while the target population is that of children. We note that the children of a particular family can be selected through either the father or the mother. Figure 1 provides a simple illustration for this indirect sampling scheme (Figure 1.2, Lavallée 2007).

There is a sizeable amount of literature concerning estimation problems that are associated with indirect sampling, a few of which we name here. Initially, estimation methods for production of cross-sectional estimates using longitudinal household survey are discussed in Ernst (1989). This study presents weight share method in the context of longitudinal survey and also shows that this method provides an unbiased estimator for the total for any characteristic in the population of interest. Kalton and Brick (1995) conclude that such a method also provides minimal variance of estimated population total for some simple sampling schemes for the longitudinal household panel survey. Lavallée (1995) extends weight share method in a completely general context of indirect sampling which includes longitudinal survey as its particular example, called generalized weight share method (GWSM). This work justifies that this weighting scheme provides unbiased estimates irrespective of sampling schemes in obtaining a sample in the sampling population. As with any other weighting scheme, in the process of GWSM implementation an adjustment for a variety of nonresponse problems must be made. Lavallée (2001) provides adjusted GWSM incorporating possible total nonresponse problems in indirect sampling. In indirect sampling there is another type of nonresponse called link nonresponse, termed by Lavallée (2001) as "relationship nonresponse," which is associated with a situation where it is impossible to determine, or where one has failed to determine, whether or not a unit in the sampling population is related to a unit in the target population. Lavallée (2001) points out the problem of overestimation in using GWSM when link nonresponse occurs and leaves finding suitable adjustment of GWSM for link nonresponse as a rather open question. This present study focuses on developing treatments of estimation bias caused by such link nonresponse.

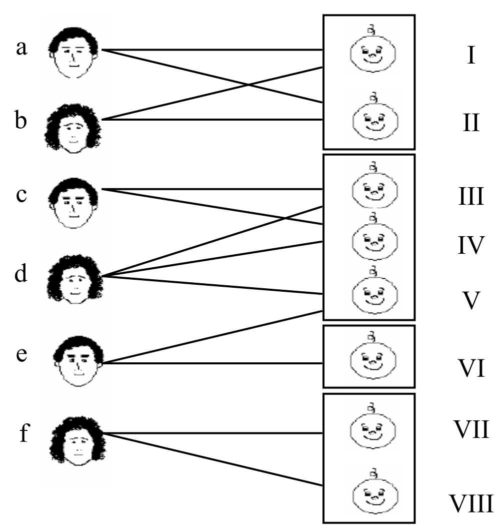

Figure 1 Indirect sampling of children

1. Xiaojian Xu, Department of Mathematics, Brock University, St. Catharines, Ontario, Canada, L2S 3A1. E-mail: xxu@brocku.ca; Pierre Lavallée, Social Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6. E-mail: Pierre.lavallée@statcan.gc.ca.

The rest of this work has been arranged in the following sections. Notation and the problem defined are described in Section 2. We propose a few modification methods in using GWSM incorporating link nonresponse in Section 3. A simulation study using a real life data set is presented in Section 4 with a few closing remarks in Section 5. We note that we show the advances of the new methods provided in this paper through a simulation study while other theoretical contributions relevant to this problem can be found in Lavallée (2002), Deville and Lavallée (2006), and Lavallée (2007).

2. Notation and problem

We use U^A and U^B to denote sampling population and target population respectively. Then, U^A is the population related to U^B with a known sampling frame. We let s^A, M^A, and m^A be a selected sample from U^A, the number of units in U^A, and the number of units in s^A respectively. We use π_j^A to represent the selection probability of j^{th} unit in U^A with $\pi_j^A > 0$ and $\sum_{j=1}^{M^A} \pi_j^A = m^A$. We also make use of the notation: M^B, N, U_i^B, and M_i^B to be the number of units in U^B, the number of clusters in U^B, the i^{th} cluster of U^B with $\cup_{i=1}^N U_i^B = U^B$, and the number of units in i^{th} cluster U_i^B.

We define $l_{j,ik}$ as an indicator variable of link existence: $l_{j,ik} = 1$ indicates that there is a link between j^{th} unit in U^A and k^{th} unit in U_i^B, while $l_{j,ik} = 0$ indicates otherwise. We also define $L_{j,i}^B$ as the total number of links existing between unit j of U^A and units of U_i^B, *i.e.*, $L_{j,i}^B = \sum_{k=1}^{M_i^B} l_{j,ik}$. Let L_i^B be the total number of links existing between units of U^A and units of U_i^B, *i.e.*, $L_i^B = \sum_{j=1}^{M^A} L_{j,i}^B$. We denote the value of the characteristics for the k^{th} unit of i^{th} cluster in population U^B by y_{ik}, and the total of all $y'_{ik}s$ by Y^B. Then, we have $Y^B = \sum_{i=1}^N \sum_{k=1}^{M_i^B} y_{ik}$.

We let Ω^B denote the clusters in U^B where there is at least one unit ik such that $l_{j,ik} = 1$ for some j^{th} unit in s^A, and we say that it can be identified by units j in s^A, *i.e.*, such i satisfies $L_i^B = \sum_{j=1}^{M^A} \sum_{k=1}^{M_i^B} l_{j,ik} > 0$. The number of clusters in Ω^B is n. After sampling we relabeled the clusters in Ω^B as $i = 1, 2, \ldots, n$. We let w_{ik} refer to the estimation weight assigned to k^{th} unit of i^{th} cluster, Ω_i^A refer to the set of units in U^A that have links to some units in U_i^B with $i \in \Omega^B$, and Ω^A refer to the set of units in U^A that have links to some units in Ω^B, *i.e.*, $\Omega^A = \{j \mid \sum_{i \in \Omega^B} L_{j,i}^B \neq 0\}$. We use s_i^A to indicate the set of units in s^A that have links to some units in U_i^B with $i \in \Omega^B$. We let T^A, T_i^A, and m_i^A denote the number of units in Ω^A, the number of units in Ω_i^A, and the number of units in s_i^A respectively. Finally, we make use of the following three indicators: let t_j be the indicator variable of being selected in $s^A: t_j = 1$ indicates that j^{th} unit in U^A is in s^A and $t_j = 0$ indicates otherwise; let t_j^L be the indicator variable of being included in s^A for units in $\Omega^A: t_j^L = 1$ indicates that j^{th} unit in Ω^A is in s^A and $t_j^L = 0$ indicates otherwise; and let $t_{j,i}^L$ be the indicator variable of being included in s_i^A for units in $\Omega_i^A: t_{j,i}^L = 1$ indicates that j^{th} unit in Ω_i^A is in s_i^A and $t_{j,i}^L = 1$ indicates otherwise.

Our goal is to estimate the total Y^B, the parameter of our interest, for target population U^B which is divided into N clusters. In order to do so, we select a sample s^A from U^A with selection probability π_j^A. Then we identify Ω^B using $l_{j,ik} \neq 0$. All units of the clusters in Ω^B are surveyed where y_{ik} and the set of $l_{j,ik}$ are measured.

By applying the GWSM, an estimation weight w_{ik} will be assigned to each unit k of surveyed cluster i's. Such weights can be chosen in an appropriate manner so that the estimator of Y^B:

$$\hat{Y}^B = \sum_{i=1}^{n} \sum_{k=1}^{M_i^B} w_{ik} y_{ik} \tag{1}$$

performs well in estimating Y^B.

We are interested in estimating the quantity Y^B using $\hat{Y}^B$. According to Horvitz and Thompson (1952), let w_{ik} be inverse of selection probability, π_{ik}, of the k^{th} individual of U_i^B in the target population. Then $\hat{Y}^B$ gives an unbiased estimator for Y^B. However, the computation for π_{ik} is difficult or even impossible in the present case, due to the complication in the indirect sampling scheme. Therefore, GWSM is introduced to address this issue. For readers' convenience, here we outline the GWSM in computing the weights for each cluster that has been observed.

Step 1: Provide the initial weights w'_{ik}

$$w'_{ik} = \sum_{j=1}^{M^A} l_{j,ik} \frac{t_j}{\pi_j^A}; \tag{2}$$

Step 2: Compute L_i^B

$$L_i^B = \sum_{k=1}^{M_i^B} \sum_{j=1}^{M^A} l_{j,ik}; \tag{3}$$

Step 3: Obtain final weight w_i

$$w_i = \frac{\sum_{k=1}^{M_i^B} w'_{ik}}{L_i^B}; \tag{4}$$

Step 4: Set $w_{ik} = w_i$ for all k in i^{th} cluster.

It follows Theorem in Section 3 of Lavallée (2001) that

$$\hat{Y}^B = \sum_{i=1}^{n} \frac{\sum_{j=1}^{M^A} L_{j,i}^B \frac{t_j}{\pi_j^A}}{L_i^B} \sum_{k=1}^{M_i^B} y_{ik} \tag{5}$$

offers an unbiased estimator for Y^B provided all links $l_{j,ik}$ can be correctly identified. The estimation weights assigned in (5) are

$$w_{ik} = \begin{cases} \dfrac{\sum_{j=1}^{M^A} L_{j,i}^B \frac{t_j}{\pi_j^A}}{L_i^B}, & \text{for all units } k \text{ in cluster } i \text{ when } i \text{ in } \Omega^B; \\ 0, & \text{when } i \text{ is not in } \Omega^B. \end{cases} \tag{6}$$

A simple example is illustrated in Figure 2. We aim to estimate the total Y^B linked to the target population U^B. Suppose that we select the units $j = 1$, and 2 from U^A. By selecting the unit $j = 1$, we survey the units of cluster $i = 1$. Likewise, by selecting the unit $j = 2$, we survey the units of clusters $i = 1$, and 2. We therefore have $\Omega^B = \{1, 2\}$. For each unit k of clusters i of Ω^B, we calculate the initial weights w'_{ik} in (2), the total number of links existing between units of U^A and units of U_i^B, L_i^B, and the final weights w_{ik}. Then, according to (5) the resulting estimator for Y^B is as below (see Lavallée 2007, pages 17-18 for more details):

$$\hat{Y}^B = \frac{1}{2}\left[\frac{1}{\pi_1^A} + \frac{1}{\pi_2^A}\right] y_{11} + \frac{1}{2}\left[\frac{1}{\pi_1^A} + \frac{1}{\pi_2^A}\right] y_{12} + \frac{1}{3\pi_2^A} y_{21} + \frac{1}{3\pi_2^A} y_{22} + \frac{1}{3\pi_2^A} y_{23}. \tag{7}$$

We note that for the estimator with known $l_{j,ik}$, the only assumption for unbiasedness is to have $L_i^B > 0$ for all clusters i's in U^B. That is, every cluster of the target population must have at least one link from U^A. We know that if some links were missing, then the estimator (5) would be biased. When link nonresponse occurs, as indicated in Lavallée (2001), L_i^B can not be determined. Traditionally, using total links observed to replace this unknown quantity results in overestimation on Y^B since some link components are actually missing in summation L_i^B. Our proposed study focus is on just such a problem, and we attempt to adjust the estimation weights w_{ik} by estimating L_i^B so as to obtain a better performance of estimation on Y^B.

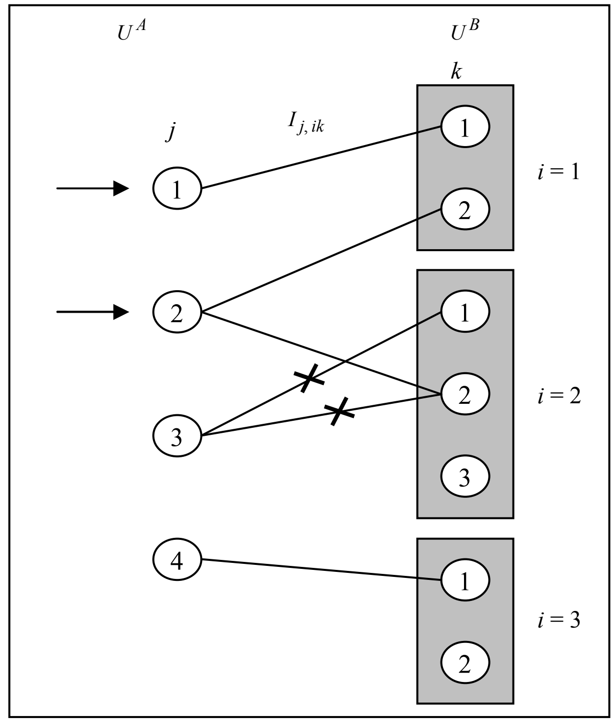

Figure 2 Example of links in indirect sampling

3. Treatments of biased estimation problems

As indicated in Section 1, the biased estimation using GWSM occurs due to link nonresponse problems. In this situation, not all of the composition in L_i^B can be identified or observed. Although the links between units in s^A and units in U^B can normally be determined in practice, the parts of links outside s^A are often difficult or even impossible to identify. We say that such units have missing links with U^B. Let $\Delta^A = \Omega^A \setminus s^A$ be the set of units with possible missing links. Then,

$$L_i^B = \sum_{j \in s^A} \sum_{k=1}^{M_i^B} l_{j,ik} + \sum_{j \in \Delta^A} \sum_{k=1}^{M_i^B} l_{j,ik}. \tag{8}$$

If we carry out the GWSM without taking these missing links into account, we use the total of observed $l_{j,ik}$ as L_i^{B*} instead to compute $\hat{Y}^B$ using

$$L_i^{B*} = \sum_{j \in s^A} \sum_{k=1}^{M_i^B} l_{j,ik} + \sum_{j \in \Delta_0^A} \sum_{k=1}^{M_i^B} l_{j,ik}, \tag{9}$$

where Δ_0^A is a subset of Δ^A and only contains the units whose links are observed. The cost is overestimation of Y^B in using (5) since

$$L_i^B \geq L_i^{B*}.$$

We suggest a few methods for applying GWSM under consideration of link nonresponse by estimating L_i^B.

3.1 Estimating L_i^B without availability of auxiliary variables

3.1.1 Estimating L_i^B by proportional adjustment for each individual cluster (Method 1)

To address the link nonresponse problem, we focus on estimating L_i^B using the known information about the links within s^A. To compute the weights in (6) using GWSM, we only need to estimate L_i^B for those $i \in \Omega^B$. For any $i \in \Omega^B$,

$$L_i^B = \sum_{j=1}^{T_i^A} L_{j,i}^B. \tag{10}$$

A general estimator for this total can be expressed as

$$\hat{L}_i^B = \sum_{j=1}^{T_i^A} w_{j,i}^L L_{j,i}^B, \tag{11}$$

where $w_{j,i}^L$ is a random weight that takes the value $w_{j,i}^L = 0$ if j is not in the sample s_i^A. For each $i \in \Omega^B$, we use the known link information between s_i^A and U_i^B to estimate the link information between Ω_i^A and U_i^B. The expectation of $\hat{L}_i^B$ is

$$E(\hat{L}_i^B) = \sum_{j=1}^{T_i^A} E(w_{j,i}^L) L_{j,i}^B. \tag{12}$$

By comparing (10) and (12), it can be observed that $\hat{L}_i^B$ is unbiased for L_i^B for any weighting scheme with $E(w_{j,i}^L) = 1$ for all j.

First of all, we adopt the Horvitz-Thompson estimator (Horvitz & Thompson 1952), also called π estimator (Särndal, Swensson, and Wretman 1991). Note that, by the definition of $\Omega_i^A, \Omega_i^A \supset s_i^A$ for all i. We *imitate* a procedure for estimating the number of links in Ω_i^A using that in s_i^A. The procedure is to select a "sample" s_i^A from the "population" Ω_i^A. Let $\pi_{j,i}^L$ be the probability of j (which is in Ω_i^A) being included in s_i^A. Then, let

$$w_{j,i}^L = \begin{cases} 1/\pi_{j,i}^L, & j \text{ is in } s_i^A, \\ 0, & j \text{ is in } \Omega_i^A \setminus s_i^A. \end{cases} \tag{13}$$

According to Corollary 3.1 in Cassel, Särndal, and Wretman (1977), this weighting scheme provides an unbiased estimator for L_i^B. We have

$$\hat{L}_i^B = \sum_{j=1}^{T_i^A} \frac{L_{j,i}^B t_j^L}{\pi_{j,i}^L}. \tag{14}$$

It provides us with an asymptotically unbiased (proof follows) estimator of Y^B:

$$\tilde{Y}^B = \sum_{i=1}^{n} \frac{\sum_{j=1}^{M^A} L_{j,i}^B \frac{t_j}{\pi_j^A}}{\sum_{j=1}^{T_i^A} \frac{L_{j,i}^B t_{j,i}^L}{\pi_{j,i}^L}} \sum_{k=1}^{M_i^B} y_{ik}. \tag{15}$$

In order to show its unbiasedness, we employ Taylor's expansion. According to Corollary 5.1.5 (Fuller 1996), we obtain

$$\begin{aligned} \frac{1}{\hat{L}_i^B} &= \frac{1}{L_i^B} - \frac{1}{(L_i^B)^2}(\hat{L}_i^B - L_i^B) + O([\hat{L}_i^B - L_i^B]^2) \\ &= \frac{1}{(L_i^B)^2}(2L_i^B - \hat{L}_i^B) + O_p(n^{-1}). \end{aligned}$$

It follows that

$$p\lim\left\{ n^{1/2}\left[\frac{1}{\hat{L}_i^B} - \frac{1}{(L_i^B)^2}(2L_i^B - \hat{L}_i^B)\right]\right\} = 0.$$

Therefore, by Theorem 5.2.1 (Fuller 1996), the limiting distribution of $n^{1/2}[1/\hat{L}_i^B]$ is the limiting distribution of $n^{1/2}[1/(L_i^B)^2(2L_i^B - \hat{L}_i^B)]$. We note that $\tilde{Y}^B$ is a function of both random variable: t_j, and random variable: $t_{j,i}^L$; therefore we denote the expectation of $\tilde{Y}^B$ with respect to t_j by $E_{t_j}(\cdot)$ and that with respect to $t_{j,i}^L$ by $E_{t_{j,i}^L}(\cdot)$. Hence, asymptotically we have

$$E(\tilde{Y}^B)$$

$$\approx \sum_{i=1}^{n} E_{t_j}\left[E_{t_{j,i}^L}\left(\frac{1}{(L_i^B)^2}\left(2L_i^B - \sum_{j=1}^{T_i^A}\frac{L_{j,i}^B t_{j,i}^L}{\pi_{j,i}^L}\right) \sum_{j=1}^{M^A} L_{j,i}^B \frac{t_j}{\pi_j^A}\right) \middle| \Omega^B \right] \sum_{k=1}^{M_i^B} y_{ik}$$

$$= \sum_{i=1}^{n} E_{t_j}\left(\frac{1}{L_i^B}\sum_{j=1}^{M^A} L_{j,i}^B \frac{t_j}{\pi_j^A}\right)\sum_{k=1}^{M_i^B} y_{ik} \tag{16}$$

$$= E_{t_j}\left(\sum_{i=1}^{n}\left(\frac{1}{L_i^B}\sum_{j=1}^{M^A} L_{j,i}^B \frac{t_j}{\pi_j^A}\right)\sum_{k=1}^{M_i^B} y_{ik}\right) \tag{17}$$

$$= E_{t_j}(\hat{Y}^B).$$

According to Lavallée (1995), $E_{t_j}(\hat{Y}^B) = Y^B$. Therefore, $\tilde{Y}^B$ is an approximately unbiased estimator of Y^B.

Now we need to compute $\pi_{j,i}^{L}$. It is a function of π_{j}^{A} yet it depends on how s_i^A affects on U_i^B, therefore on Ω_i^A. Such an effect is difficult to track and varies from case to case; however, we can give a general estimate of it. The first approach we propose in this paper is to estimate selection probability, $\pi_{j,i}^{L}$ using the proportion of the units in s^A which take in Ω^A. Namely

$$\hat{\pi}_{j,i}^{L(1)} = \frac{m_i^A}{T_i^A}. \tag{18}$$

Therefore,

$$\hat{L}_i^{B(1)} = \sum_{j=1}^{T_i^A} \frac{L_{j,i}^B\, t_j^L}{\hat{\pi}_{j,i}^{L(1)}} = \frac{T_i^A}{m_i^A} \sum_{j=1}^{m_i^A} L_{j,i}^B. \tag{19}$$

and

$$\hat{Y}^{B(1)} = \sum_{i=1}^{n} \frac{\sum_{j=1}^{M^A} L_{j,i}^B \frac{t_j}{\pi_j^A}}{\frac{T_i^A}{m_i^A}\sum_{j=1}^{m_i^A} L_{j,i}^B} \sum_{k=1}^{M_i^B} y_{ik} = \sum_{i=1}^{n} w_i^{(1)} \sum_{k=1}^{M_i^B} y_{ik}, \tag{20}$$

with

$$w_i^{(1)} = \frac{m_i^A}{T_i^A} \frac{\sum_{j=1}^{m^A} \frac{L_{j,i}^B}{\pi_j^A}}{\sum_{j=1}^{m_i^A} L_{j,i}^B}. \tag{21}$$

We revisit the example in Figure 2, assuming that there are two link nonresponses that happened between the unit $j=3$ in U^A and the units $k=1,2$ of cluster $i=2$ in U^B. If we use the GWSM without adjustment in (5), the resulting estimator for Y^B is no longer (7). We have instead

$$\hat{Y}^B = \frac{1}{2}\left(\frac{1}{\pi_1^A} + \frac{1}{\pi_2^A}\right) y_{11} + \frac{1}{2}\left(\frac{1}{\pi_1^A} + \frac{1}{\pi_2^A}\right) y_{12} + \frac{1}{\pi_2^A} y_{21} + \frac{1}{\pi_2^A} y_{22} + \frac{1}{\pi_2^A} y_{23}, \tag{22}$$

which is biased. In order to apply (20), we first compute m_i^A / T_i^A. Then the resulting weights using Method (1) in (21) for this example is shown in Table 1. Therefore, this modified method provides the estimator:

$$\hat{Y}^B = \frac{1}{2}\left(\frac{1}{\pi_1^A} + \frac{1}{\pi_2^A}\right) y_{11} + \frac{1}{2}\left(\frac{1}{\pi_1^A} + \frac{1}{\pi_2^A}\right) y_{12} + \frac{1}{2\pi_2^A} y_{21} + \frac{1}{2\pi_2^A} y_{22} + \frac{1}{2\pi_2^A} y_{23}, \tag{23}$$

which is less biased than (22).

Table 1
Initial weights, total number of responded links, and final weights from (21)

i	k	w'_{ik}	L_i^B	m_i^A	T_i^A	m_i^A/T_i^A	$w_i^{(1)}$
1	1	$1/\pi_1^A$	1	2	2	1	$1/2(1/\pi_1^A + 1/\pi_2^A)$
1	2	$1/\pi_2^A$	1	2	2	1	$1/2(1/\pi_1^A + 1/\pi_2^A)$
2	1	0	0 (missing)	1	2	1/2	$1/2\pi_2^A$
2	2	$1/\pi_2^A$	1 (one link is missing)	1	2	1/2	$1/2\pi_2^A$
2	3	0	0	1	2	1/2	$1/2\pi_2^A$

3.1.2 Estimating L_i^B by overall proportional adjustment (Method 2)

In the previous approach, the information regarding m_i^A and T_i^A is needed for every i. Suppose we ignore the variation of Ω_i^A among all i, then we simply propose that

$$L_i^{B*} = \sum_{j=1}^{T^A} \frac{L_{j,i}^B\, t_j^L}{\pi_j^L} \tag{24}$$

using link information in s^A to estimate the link information in T^A, where t_j^L being the indicator variable for being in s^A from Ω^A. Now we need to compute π_j^L. Again it is a function of π_j^A and yet it depends on the complexity of effects of s^A on Ω^B, hence to Ω^A. While the computation is difficult and varies from case to case without a general form, we can usually give a rough estimate of it.

The second approach we propose in this paper is to estimate π_j^L using the proportion of the units in s^A which appear in Ω^A, *i.e.*, $\pi_j^{L*} = m^A / T^A$. It informs us that

$$\hat{L}_i^{B(2)} = \frac{T^A}{m^A} \sum_{j=1}^{m^A} L_{j,i}^B. \tag{25}$$

For simple random designs with or without stratification, $\hat{L}_j^{B(2)}$ provides an unbiased estimator for L_i^B. For more complex designs, it provides a model-based unbiased estimator under assumption (A) as follows:

(A) Suppose that for any cluster i, the average of total existing links associated with all units in the sample s^A is the same as that of existing links associated with all units in U^A, *i.e.*,

$$\frac{\sum_{j=1}^{m^A} L_{j,i}^B}{m^A} = \frac{\sum_{j=1}^{M^A} L_{j,i}^B}{T^A}. \tag{26}$$

So, the estimation weights are provided by

$$w_{ik}^{(2)} = w_i^{(2)} = \frac{m^A}{T^A} \frac{\sum_{j=1}^{M^A} L_{j,i}^B \frac{t_j}{\pi_j^A}}{\sum_{j=1}^{M^A} L_{j,i}^B t_j}, \text{ for all units } k \text{ in cluster } i. \tag{27}$$

It follows that Y^B can be estimated by

$$\hat{Y}^{B(2)} = \frac{m^A}{T^A} \sum_{i=1}^{n} \frac{\sum_{j=1}^{m^A} \frac{L_{j,i}^B}{\pi_j^A}}{\sum_{j=1}^{m^A} L_{j,i}^B} \sum_{k=1}^{M_i^B} y_{ik} = \sum_{i=1}^{n} w_i^{(2)} \sum_{k=1}^{M_i^B} y_{ik}, \tag{28}$$

We recall the example in Figure 2 with two link nonresponses that happened between the unit $j = 3$ in U^A and the units $k = 1, 2$ of cluster $i = 2$ in U^B. In order to apply (28), we first compute m^A / T^A. For this example, we have $m^A = 2$, and $T^A = 3$. Then the resulting estimator for Y^B using the adjustment Method (2) for this example is

$$\hat{Y}^B = \frac{2}{3}\left[\frac{1}{2}\left(\frac{1}{\pi_1^A} + \frac{1}{\pi_2^A}\right) y_{11} + \frac{1}{2}\left(\frac{1}{\pi_1^A} + \frac{1}{\pi_2^A}\right) y_{12} + \frac{1}{\pi_2^A} y_{21} + \frac{1}{\pi_2^A} y_{22} + \frac{1}{\pi_2^A} y_{23}\right]. \tag{29}$$

Therefore, this adjustment made in (28) is different from Method (1) for this example.

We know that $\text{var}(\hat{Y}^{B(1 \text{ or } 2)}) = \text{var}\{E(\hat{Y}^{B(1 \text{ or } 2)} \mid s^A)\} + E\{\text{var}(\hat{Y}^{B(1 \text{ or } 2)} \mid s^A)\}$. The inner expectation and variance (conditional on s^A) are taken over all possible sets of "responding" $l_{j,ik}$, given the sample s^A while the outer expectation and variance are taken over all possible sample s^A. Generally, the adjustments made above will not eliminate the second term which depends on the randomness of $l_{j,ik}$.

3.2 Estimating L_i^B with availability of auxiliary variables

3.2.1 Estimating $l_{j,ik}$ using logistic model

The estimation methods for L_i^B proposed in Section 3.1 are simple to apply and do not need additional information. However, sometimes the assumption can be violated which results in an undesirable estimate. For instance, $L_{j,i}^B$ may depend on some characteristics of unit *j* and cluster *i*.

We assume that the probability of a link between a unit in sampling population and a unit in target population depends on some auxiliary variables through a logistic regression model. We may estimate this probability function so that the estimation of the quantity of interest in the target population is desirable. Let $P_{j,ik} = P(l_{j,ik} = 1)$ which is affected by some variable vector $\mathbf{x}_j^A$ in U^A and $\mathbf{x}_{ik}^B$ in U^B.

We may fit the logistic model

$$\log\left(\frac{P_{j,ik}}{1 - P_{j,ik}}\right) = \mathbf{a}'\mathbf{x}_j^A + \mathbf{b}'\mathbf{x}_{ik}^B \tag{30}$$

using the observed links and their corresponding characteristic variables. The unknown parameter vectors $\mathbf{a}$ and $\mathbf{b}$ can be estimated. Then, for those $l'_{j,ik}s$ which can not be identified we suggest to impute them with their probability estimates:

$$\hat{P}_{j,ik} = \frac{e^{\hat{\mathbf{a}}'\mathbf{x}_j^A + \hat{\mathbf{b}}'\mathbf{x}_{ik}^B}}{1 + e^{\hat{\mathbf{a}}'\mathbf{x}_j^A + \hat{\mathbf{b}}'\mathbf{x}_{ik}^B}}, \tag{31}$$

where $(\hat{\mathbf{a}}, \hat{\mathbf{b}})$ is an estimator for $(\mathbf{a}, \mathbf{b})$, for instance, we use the weighted maximum likelihood (pseudolikelihood) estimator. We then have

$$\begin{aligned} \hat{L}_i^{B(3)} &= \sum_{j \in s^A \cup \Delta_0^A} L_{j,i} + \sum_{j \in \Omega^A \backslash (s^A \cup \Delta_0^A)} \hat{L}_{j,i} \\ &= \sum_{j \in s^A \cup \Delta_0^A} L_{j,i} + \sum_{j \in \Omega^A \backslash (s^A \cup \Delta_0^A)} \sum_{k=1}^{M_i^B} \frac{e^{\hat{\mathbf{a}}'\mathbf{x}_j^A + \hat{\mathbf{b}}'\mathbf{x}_{ik}^B}}{1 + e^{\hat{\mathbf{a}}'\mathbf{x}_j^A + \hat{\mathbf{b}}'\mathbf{x}_{ik}^B}}. \end{aligned} \tag{32}$$

After replacing L_i^B with $\hat{L}_i^{B(3)}$ in (5), (5) provides us with a consistent estimator for Y^B when the model specified in (30) is correct and $(\hat{\mathbf{a}}, \hat{\mathbf{b}})$ is consistent. Note that there are alternatives for the logistic model, such as logit and complementary log-log models. See Draper and Smith (1998) for details. Their research also states that the choice of which model should be employed is not always clear in practice.

3.2.2 Directly estimating L_i^B use log-linear model

We consider that there is a variable vector $\mathbf{x}_i^B$ which affects the value of L_i^B. This indicates that the total number of links in a cluster only varies according to the characteristics of the cluster itself. Using the log-linear model, we can propose (33) below:

$$\log(L_i^B) = \theta^T \mathbf{x}_i^B. \tag{33}$$

If the fit is reasonable, L_i^B can be estimated directly by

$$\hat{L}_i^{B(4)} = e^{\hat{\theta}^T \mathbf{x}_i^B}, \tag{34}$$

where $\hat{\theta}$ is an estimator for θ. When $\hat{\theta}$ is consistent then after replacing L_i^B with $\hat{L}_i^{B(4)}$ in (5), (5) provides a consistent estimator for Y^B. We note that $\hat{L}_i^{B(4)}$ might be non-integer valued, and therefore might have to be rounded to the nearest integer value.

4. Simulation study

When the production of cross-sectional estimates at a particular point in time after the initial point is also of interest in a longitudinal survey design, it becomes a practical example of an indirect sampling problem. Since the population changes over time, the target population is not the same as the initial population which the longitudinal sample is selected from. In this section we will use Survey of Labour and Income Dynamics (SLID) as an example to demonstrate the performance of one of the estimators we introduced in Section 3.1.

The sample design for SLID is detailed in Lavallée (1993). Some terminologies we use in this report - such as cohabitants, initially-present individuals, and initially-absent individuals - follow Lavallée (1995). Initially-absent individuals in the population are individuals who were not part of the population in the year the longitudinal sample was selected, but are considered in the later sample; included among these are newborns and immigrants. After the initial year of selection, the population contains longitudinal individuals, initially-present individuals and initially-absent individuals. Focusing on the households containing at least one longitudinal individual (*i.e.*, longitudinal households), initially-present and initially-absent individuals who join these households are referred to as cohabitants.

In this specific example, U^A is the population at the initial year, say yr_0, of the longitudinal survey, and U^B is the population at any of the following years, say year yr_t, after the initial year. The sample s^A is all the longitudinal individuals. $L_{j,i}$ is a binary variable; it values 1 if individual j lives in i^{th} household at yr_t; 0 otherwise. L_i^B is the total number of longitudinal persons and initially-present cohabitants at yr_0 who lives in i^{th} household at yr_t.

For a longitudinal individual the link would be one to one. For cohabitants there is a significant possibility that this link will be impossible to identify a few years past the initial year, for reasons such as new birth and immigration; further, the greater proportion of cohabitants occupying the target population, the larger this possibility becomes. For instance, in survey panel 3 in SLID, cohabitants represent 7.8 percents out of 47,377 individuals in the year of 2000 which is one year after the initial year. This increases to 13.87 percent in the year 2002 (3 years later), and 15.22 percent in 2003 (4 years later). We can see that the link nonresponses can not be overlooked in such a significant proportion of cohabitants. Due to the availability of observed information, we implement the approach of estimating L_i^B by two kinds of proportional adjustments, which we proposed in Section 3.1.1 and 3.1.2. In order to test the performance of the estimates obtained by these approaches, we carry out a simulation study using SLID data. Cross-sectional estimations for four income variables are of interest for the year of 2003. These four variables are: total income before taxes; total income after taxes; earnings (includes wages and salaries before deductions and self-employment income); and wages and salaries before deductions (also called employment income). We are interested in the total of the population incomes for all these variables. These four quantities of interest have been estimated at both the national level and the provincial level.

For a longitudinal survey, the total number of links in cluster i are generally not more than the total number of individuals in this cluster and not less than the number of longitudinal individuals in this cluster. Since T_i^B is unknown, we replace T_i^B by M_i^B in (5) in our simulation study.

First, we assume that the links between all units selected in the initial year (1999) and all units in the whole population in 2003 are correctly specified. Then we compute the totals using GWSM. We use it as our estimation target, the "truth."

Second, we randomly take away 50 percent of the links associated with initially-present individuals by setting up at random some initially present cohabitants as initially absent ones. The number of links taken makes up approximately 6.3 percent of the total population with which we are interested, with a size of 30,224. Without any adjustment, we recalculate the estimates using GWSM. We use it as our estimation benchmark, the "placebo."

Third, we estimate the same quantities using GWSM with proportional adjustment approaches, Method (1) and (2) in Section 3.1, to see whether the estimates are close enough to the "truth" and how much improvement these adjustments make.

This simulation study using SLID data demonstrates that the proposed method performs very well in overcoming the overestimation problems that arise from link nonresponse.

We denote

$$w_i^{\text{mean}} = \frac{\sum_{j=1}^{m^A} L_{j,i}^B \frac{1}{\pi_j^A}}{\sum_{j=1}^{m^A} L_{j,i}^B} \tag{35}$$

Then, using Method (1) and (2) in Section 3.1 we estimate Y^B by

$$\hat{Y}_{\text{mean}}^{B(1)} = \sum_{i=1}^{n} \frac{m_i^A}{T_i^A} w_i^{\text{mean}} \sum_{k=1}^{M_i^B} y_{ik}, \tag{36}$$

and

$$\hat{Y}^{B(2)}_{\text{mean}} = \frac{m^A}{T^A} \sum_{i=1}^{n} w_i^{\text{mean}} \sum_{k=1}^{M_i^B} y_{ik}, \qquad (37)$$

respectively.

We note that w_i^{mean} is the average weight of longitudinal persons who live in i^{th} household at yr_t. Therefore, it is also reasonable to use median weight:

$$w_i^{\text{median}} = \text{the median of } \frac{1}{\pi_j^A}, j = 1, 2, \ldots, m^A. \qquad (38)$$

instead to enhance the robustness of the estimates. Namely, we estimate Y^B as well by

$$\hat{Y}^{B(1)}_{\text{median}} = \sum_{i=1}^{n} \frac{m_i^A}{T_i^A} w_i^{\text{median}} \sum_{k=1}^{M_i^B} y_{ik}, \qquad (39)$$

and

$$\hat{Y}^{B(2)}_{\text{median}} = \frac{m^A}{T^A} \sum_{i=1}^{n} w_i^{\text{median}} \sum_{k=1}^{M_i^B} y_{ik}. \qquad (40)$$

The comparison for these proposed methods with and without incorporation in nonresponse problems both using mean and median weight within each household are presented in Tables 2-5.

The next four tables give the result for the performance of our estimate using relative error defined as:

$$\left| \frac{\text{estimate - “truth”}}{\text{“truth”}} \right| \times 100\%.$$

Table 2
Total income before taxes (in Canadian dollars)

Province	Estimates by GWSM without missing links	Estimates by GWSM with missing links	Estimates by adjusted GWSM using mean	Estimates by adjusted GWSM using median
NFL	9,261,958,108	9,788,749,735	9,317,420,236	9,304,530,248
PEI	2,720,448,008	2,858,506,466	2,735,943,043	2,734,922,451
NS	18,277,017,251	19,573,546,299	18,140,076,618	18,067,144,557
NB	15,297,155,323	16,281,178,934	15,291,696,585	15,236,482,035
QC	1.57839E+11	1.69664E+11	1.56533E+11	1.56405E+11
ON	2.895E+11	3.07642E+11	2.85409E+11	2.85599E+11
MA	23,436,397,548	25,043,168,032	23,632,717,226	23,553,543,216
SK	20,185,285,649	21,595,804,296	20,163,683,598	20,095,359,071
AB	69,063,402,292	74,576,351,600	68,716,661,193	68,582,541,733
BC	81,749,374,346	86,593,614,506	81,387,640,982	81,248,680,715
National	6.8733E+11	7.33617E+11	6.8286E+11	6.82356E+11

Table 3
Total income after taxes (in Canadian dollars)

Province	Estimates by GWSM without missing links	Estimates by GWSM with missing links	Estimates by adjusted GWSM using mean	Estimates by adjusted GWSM using median
NFL	7,846,587,557	8,287,351,908	7,892,754,014	7,882,437,105
PEI	2,300,092,795	2,416,503,441	2,314,256,124	2,313,544,320
NS	15,154,508,564	16,257,679,161	15,080,155,194	15,020,088,623
NB	12,878,350,198	13,718,260,686	12,894,700,593	12,849,252,205
QC	1.27632E+11	1.37514E+11	1.27118E+11	1.26999E+11
ON	2.3788E+11	2.53073E+11	2.35192E+11	2.3534E+11
MA	19,541,510,220	20,877,377,918	19,713,628,649	19,649,142,217
SK	16,894,929,025	18,073,635,883	16,890,410,993	16,834,787,407
AB	57,466,974,767	62,055,315,246	57,183,814,491	57,073,904,623
BC	68,710,569,670	72,770,595,462	68,431,531,373	68,309,055,749
National	5.66306E+11	6.05044E+11	5.63958E+11	5.63518E+11

Table 4
Earnings (in Canadian dollars)

Province	Estimates by GWSM without missing links	Estimates by GWSM with missing links	Estimates by adjusted GWSM using mean	Estimates by adjusted GWSM using median
NFL	6,433,112,169	6,837,522,157	6,541,306,193	6,530,174,122
PEI	1,898,192,704	2,019,341,995	1,964,066,449	1,962,669,664
NS	12,772,667,160	13,809,197,160	12,999,111,234	12,939,785,579
NB	11,250,688,811	12,030,378,710	11,411,530,716	11,370,222,533
QC	1.18878E+11	1.28949E+11	1.19797E+11	1.19717E+11
ON	2.27577E+11	2.43404E+11	2.26812E+11	2.27092E+11
MA	17,560,695,670	18,995,682,322	18,066,353,153	18,001,882,362
SK	15,159,319,031	16,340,668,148	15,381,733,004	15,319,210,228
AB	56,152,023,359	61,059,244,608	56,540,145,524	56,418,889,147
BC	60,532,655,979	64,499,398,960	61,192,920,832	61,085,986,951
National	5.28214E+11	5.67945E+11	5.3199E+11	5.31722E+11

Table 5
Wages and salaries before deductions (in Canadian dollars)

Province	Estimates by GWSM without missing links	Estimates by GWSM with missing links	Estimates by adjusted GWSM using mean	Estimates by adjusted GWSM using median
NFL	6,180,713,343	6,572,345,010	6,283,079,555	6,272,429,515
PEI	1,636,344,440	1,747,755,878	1,713,809,312	1,713,157,676
NS	12,327,220,137	13,341,912,666	12,579,519,733	12,521,159,025
NB	10,742,381,379	11,508,445,078	10,961,105,589	10,921,102,477
QC	1.08636E+11	1.18092E+11	1.10024E+11	1.09898E+11
ON	2.07331E+11	2.22043E+11	2.07265E+11	2.07495E+11
MA	16,146,993,217	17,504,024,442	16,701,823,718	16,641,840,086
SK	13,982,423,360	15,129,217,320	14,311,467,435	14,255,519,224
AB	52,594,490,290	57,359,188,114	53,195,227,508	53,077,388,907
BC	56,206,787,033	59,886,429,369	56,875,663,895	56,764,297,512
National	4.85784E+11	5.23184E+11	4.91116E+11	4.90763E+11

Table 6
Comparison of relative errors in estimating income before taxes (%)

Province	GWSM with missing links	Method (1) using mean	Method (1) using median	Method (2) using mean	Method (2) using median
NFL	5.688	0.599	0.460	1.059	2.397
PEI	5.075	0.570	0.532	2.859	4.063
NS	7.094	0.749	1.148	3.549	2.459
NB	6.433	0.037	0.397	2.693	2.987
QC	7.492	0.828	0.909	4.372	2.896
ON	6.267	1.413	1.348	4.691	1.771
MA	6.856	0.838	0.500	1.644	3.654
SK	6.988	0.107	0.446	2.480	2.598
AB	7.982	0.502	0.696	3.185	2.407
BC	5.926	0.442	0.612	3.995	3.343
National	6.734	0.650	0.724	3.868	2.662

Table 7
Comparison of relative errors in estimating income after taxes (%)

Province	GWSM with missing links	Method (1) using mean	Method (1) using median	Method (2) using mean	Method (2) using median
NFL	5.617	0.588	0.457	1.101	2.409
PEI	5.061	0.616	0.585	2.832	4.121
NS	7.279	0.491	0.887	3.338	2.765
NB	6.522	0.127	0.226	2.539	3.150
QC	7.742	0.403	0.496	3.991	3.375
ON	6.387	1.130	1.068	4.432	2.081
MA	6.836	0.881	0.551	1.645	3.733
SK	6.977	0.027	0.356	2.406	2.675
AB	7.984	0.493	0.684	3.180	2.415
BC	5.909	0.406	0.584	3.989	3.419
National	6.841	0.415	0.492	3.657	2.927

Table 8
Comparison of relative errors in estimating earnings (%)

Province	GWSM with missing links	Method (1) using mean	Method (1) using median	Method (2) using mean	Method (2) using median
NFL	6.286	1.682	1.509	0.041	3.585
PEI	6.382	3.470	3.397	0.0739	7.115
NS	8.115	1.773	1.308	1.265	5.281
NB	6.930	1.430	1.062	1.279	4.512
QC	8.472	0.773	0.706	2.827	4.560
ON	6.955	0.336	0.213	3.760	2.920
MA	8.172	2.879	2.512	0.291	5.835
SK	7.793	1.467	1.055	0.979	4.324
AB	8.739	0.691	0.475	2.140	3.777
BC	6.553	1.091	0.914	2.643	5.081
National	7.522	0.715	0.664	2.628	4.131

They show that our estimates using both method (1) and method (2) perform very well in terms of reducing bias. Method (1) does work better than Method (2) overall, yet the improvement from Method (1) to Method (2) is much less compared to that made by moving from without adjustment to method (2). Since Method (2) provides us with high quality and involves much less information than Method (1), Method (2) is recommended.

Now, we focus on Method (2) using mean, which gives the estimate $\hat{Y}^{B(2)}_{\text{mean}}$, to analyze how its variance performs in terms of estimating Y^B. We use the bootstrap technique to estimate the variance of $\hat{Y}^{B(2)}_{\text{mean}}$ at both the national level and the provincial level. The bootstrap used for our simulation in this paper is the classical Bootstrap with replacement, where bootstrapping is performed at the first stage of sampling. The bootstrap weights taken here are provided with the SLID data, and incorporate all the necessary adjustments. See Lévesque (2001), and LaRoche (2003) for details on the use of the Bootstrap for SLID. The improvement in reducing the variance is not as large as in reducing bias; however, it is revealed in this simulation study that the proposed method provides a smaller variance as well compared to applying GWSM without an adjustment for missing links. See Table 10 for the results.

The simulation results presented here are based on a single sample of SLID and a single random removal of the links of initially-present individuals. For a complete assessment of the properties of the above estimators, a Monte-Carlo process would have been suitable. Such simulations have been performed by Hurand (2006) based on agricultural data. In these simulations, 1,000 samples have been selected and for each selected sample, the worst-case-scenario has been used, *i.e.*, all links from the non-sample units have been removed. The results of these simulations showed that proportional adjustment and global proportional adjustment are the two methods whose estimates are, on average, the closest to the real total, and whose biases are negligible.

Table 9
Comparison of relative errors in estimating wages and salaries before deductions (%)

Province	GWSM with missing links	Method (1) using mean	Method (1) using median	Method (2) using mean	Method (2) using median
NFL	6.336	1.656	1.484	0.1012	3.593
PEI	6.809	4.734	4.694	1.056	8.424
NS	8.231	2.047	1.573	0.939	5.509
NB	7.131	2.036	1.664	0.685	5.133
QC	8.704	1.278	1.162	2.294	5.070
ON	7.096	0.0317	0.0791	3.473	3.265
MA	8.404	3.436	3.065	0.787	6.469
SK	8.202	2.353	1.953	0.107	5.213
AB	9.059	1.142	0.918	1.713	4.247
BC	6.547	1.190	0.992	2.565	5.234
National	7.699	1.098	1.025	2.251	4.541

Table 10
Comparison of standard deviation estimates

	Variables	Total income before taxes	Total income after taxes	Earnings	Wages and salaries before deductions
National	GWSM with missing links	9,677,258,789	7,343,792,762	8,850,202,075	8,468,718,449
level	Method (2) using mean	9,471,103,083	7,238,715,323	8,593,015,854	8,232,428,642
Ontario	GWSM with missing links	7,888,106,377	6,101,001,739	7,245,688,373	7,149,203,530
	Method (2) using mean	7,601,169,501	5,939,509,894	6,952,217,872	6,831,300,511
Quebec	GWSM with missing links	4,341,215,711	3,113,247,130	3,772,369,180	3,162,277,660
	Method (2) using mean	4,160,251,472	2,974,248,451	3,668,996,929	3,100,868,366

5. Closing remarks

We have constructed four estimation methods to address the link nonresponse problem in indirect sampling. The simulation results in this article show that the adjustments methods we have presented in the example for using GWSM incorporating the link nonresponse performs well in terms of both reducing the estimation bias and providing an overall improvement in variance. The advancement in bias reduction seems significant. The implementation of the methods proposed in Section 3.2 for real data sets will be studied in the near future.

The following significant observations emerged from our study:

1. Adjustment methods are simple to apply.
2. In a more general situation, such as $L_{j,i} > 1$ for some j's, (35) represents the weighted mean weighted by $L^B_{j,i}$. Accordingly the median approach delivered by (39) and (40) can be modified using a generalized version of median – "weighted" median. Namely, we replace (38) by

$$w_i^{\text{median}} = \text{the median of } \frac{1}{\pi_j^A}$$

where $j = 1, 2, \ldots, L^B_{1,i}; 1, 2, \ldots, L^B_{2,i}; \ldots; 1, 2, \ldots, L^B_{m^A,i}$.

3. Some valid link responses outside s^A can not be used in estimating L_i^B by the methods proposed in Section 3.1. However, this valid information would be beneficial to the approaches by predicting $l_{j,ik}$ using auxiliary variables, as can be seen in Section 3.2.1.

Acknowledgements

The authors would like to thank the Associate Editor and the two referees for their helpful suggestions and comments on the previous versions of this paper. This research is funded by the Natural Sciences and Engineering Research Council of Canada, and Mathematics of Information Technology and Complex Systems.

References

Cassel, C.-M., Särndal, C.-E. and Wretman, J. (1977). *Foundations of Inference in Survey Sampling*. New York: John Wiley & Sons, Inc.

Deville, J.-C., and Lavallée, P. (2006). Indirect sampling: Foundations of the generalised weight share method. *Survey Methodology*, 32, 165-176.

Draper, N.R., and Smith, H. (1998). *Applied Regression Analysis*, 3rd Ed. New York: John Wiley & Sons, Inc.

Ernst, L. (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 & Sons, Inc., 135-159.

Horvitz, D.G., and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

Hurand, C. (2006). La méthode généralisée du partage des poids et le problème d'identification des liens. Internal report of the Social Survey Methods Division, Statistics Canada, July 2006.

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

LaRoche, S. (2003). Longitudinal and Cross-Sectional Weighting of the Survey of Labour and Income Dynamics. *Income Research Paper Series*, Catalogue no. 75F0002MIE - No. 007, Statistics Canada.

Lavallée, P. (1993). Sample representativity for the Survey of Labour and Income Dynamics. *Statistics Canada, Research Paper of the Survey of Labour and Income Dynamics*, Catalogue No. 93-19, December 1993.

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

Lavallée, P. (2001). Correcting for non-response in indirect sampling. *Proceedings of Statistics Canada's Symposium 2001*.

Lavallée, P. (2002). *Le sondage indirect*, *ou la méthode généralisée du partage des poids*. Éditions de l'Université de Bruxelles and Éditions Ellipse.

Lavallée, P. (2007). *Indirect Sampling*. New York: Springer.

Lévesque, I. (2001). Enquête sur la dynamique du travail et du revenu - Estimation de la variance. Internal document from Statistics Canada, July 2, 2001.

Särndal, C.-E., Swensson, B. and Wretman, J. (1991). *Model Assisted Survey Sampling*. New York: Springer-Verlag.