Journal of Official Statistics, Vol. 22, No. 1, 2006, pp. 35–51

Variance Estimation by Jackknife Method Under Two-Phase Complex Survey Design

Debesh Roy[1] *and Md. Safiquzzaman*[2]

The jackknife technique is applied to a general class of estimators, under two-phase sampling, for variance estimation. Considering a natural population, the performance of the weighted jackknife variance estimator has been compared with other variance estimators, including the unweighted jackknife variance estimator. A figure is included to clarify this comparison.

Key words: Auxiliary variable; generalized regression estimator; inclusion probability; jackknife estimator.

1. Introduction

The concept of the jackknife was introduced by Quenouille (1956) in connection with reduction of bias for nonlinear estimation. The possibility of using this technique for the purpose of variance estimation was proposed by Tukey (1958). Durbin (1959) may have been the first to use it in the context of finite populations. Rao (1965) and Rao and Webster (1966) consider jackknifing the classical ratio estimator. Besides the ratio estimator, many other estimators, which make use of auxiliary information are available in the literature. Three of them that are worth mentioning are the generalized regression estimator (GREG) due to Särndal (1980), the asymptotically design unbiased (ADU) estimator due to Brewer (1979), and the generalized ratio estimator due to Hajék (1971). Wright (1983) and Särndal and Wright (1984) brought all such estimators under one umbrella, called the QR-class of estimators. Roy and Safiquzzaman (2000) considered a further generalized version of the QR-class and jackknifed it to yield a fairly general jackknifed class of estimators. In that paper, jackknifing was done with the dual objective of bias reduction and variance estimation. Since the traditional jackknife procedure does not specifically adjust for the imbalance in the sample caused by varying probabilities of selection, it fails to give adequate protection to the variance estimator against selection of a sample that is highly unrepresentative of the population. In addition, since the traditional jackknife variance estimator, under unequal probability sampling, does not take into consideration the pair-wise inclusion probabilities, the quality of its performance in yielding a nominal confidence interval is diminished. This problem was alleviated by Roy and Safiquzzaman (2002) by

[1] West Bengal, India. Email: drroy@sify.com
[2] Habra High School (H.S.), Habra, North 24 Parganas, West Bengal, India. Email: safiq@sify.com
Acknowledgments: We wish to extend our thanks to two anonymous referees for their constructive criticisms and useful comments on this article, leading to improvement over an earlier draft. Our special thanks are due to the second referee for his/her constant involvement with this work. He/she examined this article with extreme care, without which the draft of it could not have taken the present shape.

introducing a weighted jackknife procedure that addresses the lack of balance caused either by varying probabilities of selection or by varying sizes of the units being selected in the sample.

Though the jackknife technique was used for inference in finite populations as early as 1959, its application to the case of two-phase sampling was not considered in the literature until Rao and Sitter (1995) examined the problem of estimation of the variance of the two-phase ratio estimator by jackknifing. Taking cues from their work, Roy and Safiquzzaman (2003) jackknifed a two-phase version of the QR-class of estimators and faced the same kind of problem in obtaining protection against a bad sample, which they had encountered earlier (2002) with the QR-class of estimators under single-phase sampling. Also their jackknifed variance estimator did not involve the pair-wise inclusion probabilities. To smooth out the imbalance in the sample caused by varying probabilities of selection and to make the variance estimator dependent on pair-wise inclusion probabilities, here we also have applied the weighted jackknife approach and have been able to derive a better jackknife variance estimator of the two-phase version of the generalized regression estimator (GREG). We compare the performance of this new jackknife variance estimator with that of the traditional one and also with that of the variance estimator based on the linearization technique. It is observed that the performance of the weighted jackknife estimator is better than that of either of its rival estimators with respect to certain chosen performance criteria. This will be demonstrated in tabular form at the end of this article.

Here the contents are divided into four parts. In Section 2, we introduce the two-phase GREG. In Section 3, we address the problem of variance estimation and arrive at an expression for a variance estimator using the linearization technique. Next we jackknife the two-phase GREG following the weighted jackknife approach and give the expression for both the weighted and the unweighted jackknife variance estimator. In Section 4, we consider a natural population of Swedish municipalities named MU284 in the book by Särndal, Swensson, and Wretman (1992), and compare the performance of the four variance estimators introduced in Section 3.

2. Notation and Two-phase GREG

Let U be a finite population on which are defined two real variables x and y, taking values x_i (>0) and y_i with totals X and Y, respectively. To estimate Y, a sample s of size n may be taken with probability $p(s)$. The design p may be assumed to admit positive inclusion probabilities π_i for unit i and π_{ij} for pairs of units (i, j) of U. By $\sum_U$ and $\sum\sum_U$, let us denote sums over i in U and i, j $(i<j)$ in U, and by $\sum_S$ and $\sum\sum_S$ those in S, respectively. Denote $Q=(Q_1, Q_2, \ldots, Q_n)'$ as a vector of parameters to be determined in terms of the design parameters and the auxiliary variables such that $Q_i>0$. The GREG estimator of Särndal (1980) in the case of a single-phase sampling plan may be written as

$$t_G=\sum_S \frac{y_i}{\pi_i}+\hat{B}_Q\left(X-\sum_S \frac{x_i}{\pi_i}\right)$$

where $\hat{B}_Q=\left(\sum_S Q_i x_i y_i\right)/\left(\sum_S Q_i x_i^2\right)$ provided x_i (>0) is known for every $i \in U$.

But when x_i's are not known for all the units of the population, this estimator is not usable. Then to estimate the unknown X we may resort to two-phase sampling in which a sample, s_1, of size n' is first drawn from U with probability $p_1(s_1)$, and then from s_1 a second phase sample, s_2, of size n is drawn with conditional probability $p_2(s_2/s_1)$. The overall two-phase sample, $s = (s_1, s_2)$, then has the selection probability $p(s) = p_1(s_1)\, p_2(s_2/s_1)$. We may denote the class of all possible first-phase samples by $\wp_1$, and that of all possible second-phase samples, for s_1 held fixed, by $\wp_2(s_1)$. For design $p_2(s_2/s_1)$, with s_1 held fixed, the inclusion probabilities

$$\pi_{2i}(s_1) = \sum_{s_2 \ni i} p_2(s_2/s_1) \quad \text{and} \quad \pi_{2ij}(s_1) = \sum_{s_2 \ni (i,j)} p_2(s_2/s_1)$$

are assumed to be positive. For fixed s_1 with $Q_i(s_1) > 0$, the two-phase version of GREG estimator may be written as

$$t'_G = \sum_{s_2} \frac{y_i}{\pi_{1i}\pi_{2i}(s_1)} + B'_Q(s_1)\left(\sum_{s_1} \frac{x_i}{\pi_{1i}} - \sum_{s_2} \frac{x_i}{\pi_{1i}\pi_{2i}(s_1)}\right)$$

where $B'_Q(s_1) = \left(\sum_{s_2} Q_i(s_1)x_i y_i\right) \Big/ \left(\sum_{s_2} Q_i(s_1)x_i^2\right)$.

It is easy to note that with the choice $Q_i(s_1) = 1/(x_i \pi_{1i} \pi_{2i}(s_1))$, t'_G reduces to the two-phase generalized ratio estimator given by

$$t'_R = \left(\sum_{s_2} \frac{y_i}{\pi_{1i}\pi_{2i}(s_1)}\right) \left[\frac{\displaystyle\sum_{s_1} \frac{x_i}{\pi_{1i}}}{\displaystyle\sum_{s_2} \frac{x_i}{\pi_{1i}\pi_{2i}(s_1)}}\right]$$

The GREG estimator t_G, introduced by Särndal (1980), was later derived by Deville and Särndal (1992) with a suitable choice of the calibration weight. Following their line of work, Roy and Safiquzzaman (2003) obtained a variant of the two-phase version of GREG estimator t'_G using a two-step calibration technique. The second term in both t_G and t'_G is the contribution of calibration that adjusts for unfortunate sample selection.

3. Variance Estimation Under Two-phase Sampling

The main objective of the present article is the estimation of the variance of the two-phase GREG estimator. This variance estimation problem was considered by Roy and Safiquzzaman (2003), applying the linearization technique and the standard jackknife technique. In that article it was shown that the performance of the standard jackknife variance estimator was better than that of the linearized variance estimator for a test set of data under two-phase sampling. However, for the reasons discussed in Section 1 and to be further discussed in the latter part of this section, the standard jackknife technique failed to give adequate protection against an unfortunate selection of an unequal probability sample. So in the present section, following the line of research pursued by Roy and Safiquzzaman (2002) in the case of single-phase sampling, the weighted jackknife variance estimation procedure has been extended to a two-phase sampling situation. In this section, we present a brief discussion of the linearized and the standard jackknife estimators, in Subsections 3.1 and 3.2, respectively. These two subsections have been drawn from an earlier work of Roy and Safiquzzaman (2002) to maintain continuity of discussion. The newest development in variance estimation by the weighted jackknife method is discussed in Subsection 3.3. Thus, in this section, we examine the problem of variance estimation by three different techniques, obtaining three different variance estimators.

3.1. Linearization Variance Estimation

Using the ordinary Taylor's series method, the approximate variance of t'_G may be derived as follows.

Denoting $e_i(s_1) = y_i - B'_Q(s_1)x_i$, approximating $B'_Q(s_1)$ by

$$\beta = E_{p_1}\left[\frac{\sum_{s_1} \pi_{2i}(s_1)Q_i(s_1)x_i y_i}{\sum_{s_1} \pi_{2i}(s_1)Q_i(s_1)x_i}\right]$$

and $e_i(s_1)$ by $E_i = E_{p_1}[e_i(s_1)]$, we can write

$$\begin{aligned}\mathrm{Var}_{p_1}E\left(t^1_G/s_1\right) \approx & \sum\sum_U \Delta_{1ij}\left(\frac{E_i}{\pi_{1i}} - \frac{E_j}{\pi_{1j}}\right)^2 + \beta^2\sum\sum_U \Delta_{1ij}\left(\frac{x_i}{\pi_{1i}} - \frac{x_j}{\pi_{1j}}\right)^2 \\ & + 2\beta\sum\sum_U \Delta_{1ij}\left(\frac{E_i}{\pi_{1i}} - \frac{E_j}{\pi_{1j}}\right)\left(\frac{x_i}{\pi_{1i}} - \frac{x_j}{\pi_{1j}}\right)\end{aligned}$$

Also

$$E_{p1}\mathrm{Var}\left(t^1_G/s_1\right) \approx \sum_{s_1\in\wp_1} p_1(s_1)\sum\sum_{s_1}\Delta_{2ij}(s_1)\left(\frac{e_i(s_1)}{\pi_{1i}\pi_{2i}(s_1)} - \frac{e_j(s_1)}{\pi_{1j}\pi_{2j}(s_1)}\right)^2$$

where $\Delta_{1ij} = \pi_{1i}\pi_{1j} - \pi_{1ij}$ and $\Delta_{2ij}(s_1) = \pi_{2i}(s_1)\pi_{2j}(s_1) - \pi_{2ij}(s_1)$.

So the linearized variance of t'_G may be written as

$$\begin{aligned}V_L(t'_G) = & \mathrm{Var}_{p_1}E\left(t^1_G/s_1\right) + E_{p_1}\mathrm{Var}\left(t^1_G/s_1\right) \approx \sum\sum_U \Delta_{1ij}\left(\frac{E_i}{\pi_{1i}} - \frac{E_j}{\pi_{1j}}\right)^2 \\ & + \beta^2\sum\sum_U \Delta_{1ij}\left(\frac{x_i}{\pi_{1i}} - \frac{x_j}{\pi_{1j}}\right)^2 + 2\beta\sum\sum_U \Delta_{1ij}\left(\frac{E_i}{\pi_{1i}} - \frac{E_j}{\pi_{1j}}\right)\left(\frac{x_i}{\pi_{1i}} - \frac{x_j}{\pi_{1j}}\right) \\ & + \sum_{s_1\in\wp_1} p_1(s_1)\sum\sum_{s_1}\Delta_{2ij}(s_1)\left(\frac{e_i(s_1)}{\pi_{1i}\pi_{2i}(s_1)} - \frac{e_j(s_1)}{\pi_{1j}\pi_{2j}(s_1)}\right)^2\end{aligned}$$

Naturally an estimate of this variance may be taken as

$$\begin{aligned}v_L(t'_G) = & \sum\sum_{s_2}\frac{\Delta_{1ij}}{\pi_{1ij}\pi_{2ij}(s_1)}\left(\frac{e_i(s_1)}{\pi_{1i}} - \frac{e_j(s_1)}{\pi_{1j}}\right)^2 + B'_Q(s_1)\sum\sum_{s_1}\frac{\Delta_{1ij}}{\pi_{1ij}}\left(\frac{x_i}{\pi_{1i}} - \frac{x_j}{\pi_{1j}}\right)^2 \\ & + 2B'_Q(s_1)\sum\sum_{s_2}\frac{\Delta_{1ij}}{\pi_{1ij}\pi_{2ij}(s_1)}\left(\frac{e_i(s_1)}{\pi_{1i}} - \frac{e_j(s_1)}{\pi_{1j}}\right)\left(\frac{x_i}{\pi_{1i}} - \frac{x_j}{\pi_{1j}}\right) \\ & + \sum\sum_{s_2}\frac{\Delta_{2ij}(s_1)}{\pi_{2ij}(s_1)}\left(\frac{e_i(s_1)}{\pi_{1i}\pi_{2i}(s_1)} - \frac{e_j(s_1)}{\pi_{1j}\pi_{2j}(s_1)}\right)^2\end{aligned}$$

3.1.1. Particular Case

With the choice $Q_i(s_1) = 1/(x_i \pi_{1i} \pi_{2i}(s_1))$, t'_G reduces to the two-phase generalized ratio estimator t'_R whose expression is given at the end of Section 2. This t'_R, in the equi-probability situation, further reduces to the two-phase ratio estimator $t'_R = \hat{R}\bar{x}'$, where $\hat{R} = \bar{y}/\bar{x}$ and $\bar{x}' = (1/n')\sum_{i=1}^{n'} x_i$. In this situation, ignoring the finite population correction, the variance estimator of t'_R takes the form

$$v_L(t'_R) \cong N^2 \left[\frac{s_e^2}{n} + \hat{R}^2 \frac{s_x'^2}{n'} + 2\hat{R} \frac{s_{ex}}{n'} \right]$$

where $s_e^2 = 1/(n-1)\sum_{s_2}(y_i - \hat{R}x_i)^2$, $s_x'^2 = 1/(n'-1)\sum_{s_1}(x_i - \bar{x})^2$ and $s_{ex} = 1/(n-1)\sum_{s_2}(y_i - \hat{R}x_i)(x_i - \bar{x})$.

Note that this $v_L(t'_R)$ is exactly the same as the linearization variance estimator by Rao and Sitter (1995) under equi-probability sampling. The calculation of $v_L(t'_R)$ has been shown in the appendix.

3.2. *Jackknife Variance Estimator*

As pointed out by Rao and Sitter (1995), the jackknife method for single phase sampling is not readily applicable to two-phase sampling. While deleting a data point for construction of pseudo-values, two cases may arise. If the jth data point being deleted belongs to s_2 then both the sample sums $\sum_{s_1}$ and $\sum_{s_2}$ in t'_G are affected, but when it belongs to $s_1 - s_2$, the only sample sum affected in t'_G is $\sum_{s_1}$. Keeping that in mind, Roy and Safiquzzaman (2003) adequately defined

$$y_R(j) = \begin{cases} \sum_{s_2} \frac{y_i}{\pi_{1i}\pi_{2i}(s_1)} - \frac{y_j}{\pi_{1j}\pi_{2j}(s_1)} & \text{if } j \in s_2 \\ \sum_{s_2} \frac{y_i}{\pi_{1i}\pi_{2i}(s_1)} & \text{if } j \in s_1 - s_2 \end{cases}$$

$$x_R(j) = \begin{cases} \sum_{s_2} \frac{x_i}{\pi_{1i}\pi_{2i}(s_1)} - \frac{x_j}{\pi_{1j}\pi_{2j}(s_1)} & \text{if } j \in s_2 \\ \sum_{s_2} \frac{x_i}{\pi_{1i}\pi_{2i}(s_1)} & \text{if } j \in s_1 - s_2 \end{cases}$$

and

$$x'_R(j) = \sum_{s_1} \frac{x_i}{\pi_{1i}} - \frac{x_j}{\pi_{1j}} \quad \text{if } j \in s_1$$

While jackknifing the estimator $t'_G (= t_{n'}$, say, when based on n' observations), let us denote by $t_{n'-1}(-j)$ the value of $t_{n'}$ based on a sample of size $(n'-1)$, obtained after having deleted the jth pair (x_j, y_j) from the original sample. Then the pseudo-values are defined as

$$t'_j = n' t_{n'} - (n'-1) t_{n'-1}(-j)$$

Here

$$t_{n'-1}(-j)=\begin{cases} y_R(j)+\dfrac{\sum_{s_2} Q_i(s_1)x_iy_i - Q_j(s_1)x_jy_j}{\sum_{s_2} Q_i(s_1)x_i^2 - Q_j(s_1)x_j^2}(x'_R(j)-x_R(j)) & \text{if } j\in s_2 \\ y_R(j)+B'_Q(s_1)(x'_R(j)-x_R(j)) & \text{if } j\in s_1-s_2 \end{cases}$$

Assuming for large sample size, $(n'-1)/n'\approx 1$, the jackknife estimator of the population total is obtained as $t_{JK}=(1/n')\sum_{i=1}^{n'} t'_i$ with jackknife variance estimator as

$$\begin{aligned} v_J(t'_G) &= \frac{1}{n'(n'-1)}\sum_{i=1}^{n'}(t'_i - t_{JK})^2 \\ &= \sum_{s_2}\left[g_{si}\frac{(y_i - B'_Q(s_1)x_i)}{\pi_{1i}\pi_{2i}(s_1)}\right]^2 + B'^2_Q(s_1)\sum_{s_1}\left(\frac{x_i}{\pi_{1i}} - \frac{1}{n'}\sum_{s_1}\frac{x_i}{\pi_{1i}}\right)^2 \\ &\quad + 2B'_Q(s_1)\sum_{s_2} g_{si}\frac{(y_i - B'_Q(s_1)x_i)}{\pi_{1i}\pi_{2i}(s_1)}\left(\frac{x_i}{\pi_{1i}} - \frac{1}{n'}\sum_{s_1}\frac{x_i}{\pi_{1i}}\right) \end{aligned}$$

where

$$g_{si} = 1 + \frac{Q_i(s_1)x_i\pi_{1i}\pi_{2i}(s_1)}{\sum_{s_1} Q_i(s_1)x_i^2}\left(\sum_{s_1}\frac{x_i}{\pi_{1i}} - \sum_{s_2}\frac{x_i}{\pi_{1i}\pi_{2i}(s_1)}\right)$$

The details of the calculation of $v_J(t'_G)$ are shown in the appendix.

3.2.1. Particular Case

With $Q_i(s_1)=1/(x_i\pi_{1i}\pi_{2i}(s_1))$, t'_G reduces under equal probability sampling to the two-phase ratio estimator $t'_R=\hat{R}\bar{x}'$; where $\hat{R}=\bar{y}/\bar{x}$ and $\bar{x}'=(1/n')\sum_{i=1}^{n'} x_i$. In this situation, the jackknife variance estimator reduces to

$$v_J(t'_R) = N^2\left[\left(\frac{\bar{x}'}{\bar{x}}\right)^2\frac{s_e^2}{n} + \hat{R}^2\frac{s'^2_x}{n'} + 2\hat{R}\left(\frac{\bar{x}'}{\bar{x}}\right)\frac{s_{ex}}{n'}\right]$$

which is exactly the same as the jackknife variance estimator of the two-phase ratio estimator proposed by Rao and Sitter (1995). The calculation of $v_J(t'_R)$ is shown in the appendix.

3.3. Weighted Jackknife Variance Estimator

Under unequal probability sampling with units having wide diversity in their sizes, the effect of deletion of units while constructing pseudo-values appears to be different for different units of the sample. In this connection, let a unit of the sample either having a large size measure or having a very high probability of inclusion, be designated as a "*heavy*" unit. Under traditional conditions, the construction of pseudo-values does not take into consideration the lack of balance caused by deletion of a "*heavy*" unit, and hence taking an unweighted average of those pseudo-values may lead to over- or under-estimation. So we proposed to take a weighted average of the present set of pseudo-values with weights fixed in such a way that while taking this average, more importance is given to the pseudo-values resulting from the deletion of a "*heavy*" unit. In fact, our weighing system absorbs the 'shock' suffered by the pseudo-values due to deletion of "*heavy*" units.

Following this line of argument, Roy and Safiquzzaman (2002), in their work on jackknife variance estimation for single-phase sampling, attached a weight function w_i to each pseudo-value t_i, and defined the weighted jackknife estimators of the population total as

$$t_{JK}(w) = \frac{\sum_S \frac{w_i}{\pi_i} t_i}{\sum_S \frac{w_i}{\pi_i}}$$

where $w_i \geq 0$ and $\sum_U w_i = 1$.

To emphasize the role of w_i as a "*shock absorber*," Roy and Safiquzzaman (2002) wrote in terms of Basu's (1971) elephant example. Borrowing words from that article, the choice of w_i can be motivated as follows. Let w_j be proportional to the current weight of the *j*th elephant in the sample. If "*Jumbo*," the heaviest elephant with small inclusion probability, is the *j*th elephant selected in the sample, then the pseudo-value resulting from the deletion of "*Jumbo*" may be written as

$$t_j = nt_n - (n-1)t_{n-1}(-j)$$

where $t_{n-1}(-j)$ is the estimate of the population total based on $s - \{Jumbo\}$.

The second part of t_j, being lighter and hence t_j being heavier than other pseudo-values, becomes prominent in the list and hence demands more attention. Thus if "*Jumbo*," the heaviest elephant with small inclusion probability, is selected against all odds, then (w_j/π_j) plays the role of an excellent buffer against the inaccuracy in $t_{n-1}(-j)$, due to deletion of "*Jumbo*."

In our present work, in an attempt to develop the concept of weighted jackknife further and to extend its application to the case of two-phase sampling, we take a weighted mean of the pseudo-values and define the weighted jackknife estimator as

$$t'_{JK}(w) = \frac{\sum_{s_2} \frac{w_i}{\pi_{1i}\pi_{2i}(s_1)} t'_i}{\sum_{s_2} \frac{w_i}{\pi_{1i}\pi_{2i}(s_1)}} \sum_{s_1} \frac{w_i}{\pi_{1i}}$$

Keeping in mind the cause of imbalance in the sample, the choice of the weight w_i, taking cues from single-phase sampling, may be determined such that

$$\text{(i)}\ \ w_i > 0 \quad \text{and} \quad \text{(ii)}\ \ E_{p_1}\left(\sum_{s_1} \frac{w_i}{\pi_{1i}}\right) = 1$$

Let us now discuss two different choices of the weights. While the first choice is designed to compensate to some degree for the imbalance caused by the varying probabilities of inclusion (as discussed by Hajék 1971), the second choice similarly compensates for the imbalance caused by the varying sizes of the units selected.

Choice 1: Here w_i's are chosen such that

$$w_i > 0, \quad w_i \infty \pi_{1i}\pi_{2i}(s_1) \quad \text{and} \quad E_{p_1}\left(\sum_{s_1}\frac{w_i}{\pi_{1i}}\right) = 1$$

An obvious choice satisfying these three conditions is

$$w_i = \frac{\pi_{1i}\pi_{2i}(s_1)}{n}$$

Choice 2: Here w_i's are chosen such that

$$w_i > 0, \quad w_i \infty X_i \quad \text{and} \quad E_{p_1}\left(\sum_{s_1}\frac{w_i}{\pi_{1i}}\right) = 1$$

The only choice here is $w_i = (X_i/X)$.

Since $t_{JK}(w)$ is a Hajék type estimator, whatever may be the choice of w_i, the weighted jackknife variance estimator is given by

$$\begin{aligned} v_{JK}(w) &= \sum\sum_{s_2}\frac{\Delta_{2ij}(s_1)}{\pi_{2ij}(s_1)}\left[\frac{w_i(t'_i - t'_{JK}(w))}{\pi_{1i}\pi_{2i}(s_1)} - \frac{w_j(t'_j - t'_{JK}(w))}{\pi_{1j}\pi_{2j}(s_1)}\right]^2 \\ &= \sum\sum_{s_2}\frac{\Delta_{2ij}(s_1)}{\pi_{2ij}(s_1)}\left(\frac{w_i t'_i}{\pi_{1i}\pi_{2i}(s_1)} - \frac{w_j t'_j}{\pi_{1j}\pi_{2j}(s_1)}\right)^2 \\ &\quad + t'^2_{JK}(w)\sum\sum_{s_2}\frac{\Delta_{2ij}(s_1)}{\pi_{2ij}(s_1)}\left(\frac{w_i}{\pi_{1i}\pi_{2i}(s_1)} - \frac{w_j}{\pi_{1j}\pi_{2j}(s_1)}\right)^2 \\ &\quad - 2t'_{JK}(w)\sum\sum_{s_2}\frac{\Delta_{2ij}(s_1)}{\pi_{2ij}(s_1)}\left(\frac{w_i t'_i}{\pi_{1i}\pi_{2i}(s_1)} - \frac{w_j t'_j}{\pi_{1j}\pi_{2j}(s_1)}\right)\left(\frac{w_i}{\pi_{1i}\pi_{2i}(s_1)} - \frac{w_j}{\pi_{1j}\pi_{2j}(s_1)}\right) \end{aligned}$$

Note: Here the objective of Choice 1 is to balance the extreme sampling weights and that of Choice 2 is to balance the extreme sample observations.

4. A Simulation Study

We consider a natural population of Swedish municipalities, named MU284 in the book by Särndal, Swensson, and Wretman (SSW) (1992). Sweden is divided into 284 municipalities having considerable variation in size and other characteristics. The data on a few variables include RMT85, the revenue from municipal taxation (in millions of Kronor) in 1985. They also include P85 and P75, the population (in thousands) in the years 1985 and 1975, respectively, for all the municipalities.

We consider the collection of these 284 municipalities as a finite population of size $N = 284$, where each municipality is considered to be a unit of the finite population. Also we take RMT85, P85 and P75 as y, x and z, respectively. From the finite population of size $N = 284$ we select R (= 10,000) first-phase samples of size n' (= 70) following the

Midzuno (1952) scheme of sampling. From each of these 10,000 first-phase samples, a second-phase sample of size n (= 20) is drawn again by employing the Midzuno (1952) scheme with z as the size measure. Then we compare the conditional performance of the linearized variance estimator and the jackknife variance estimator with that of the weighted jackknife variance estimator. R (= 10,000) samples are arranged in increasing order of magnitude of an ancillary

$$\left(\frac{\sum_{s_1} \frac{x_i}{\pi_{1i}}}{\sum_{s_2} \frac{x_i}{\pi_{1i}\pi_{2i}(s_1)}}\right)$$

that indicates the degree of balance in the sample and hence discriminates between "good" and "bad" samples. The members of the ordered set of 10,000 samples are then grouped into 10 subsets of 1,000 samples each. In each of these 10 groups, we calculate 1,000 variance estimates by the

- Linearization method (v_L),
- Jackknife method (v_J) and
- Weighted jackknife method ($v_{JK}(w) = v_{J\pi}$ with $w_i = \pi_{1i}\pi_{2i}(s_1)/n$ as weight and $v_{JK}(w) = v_{JX}$ with $w_i = x_i/X$ as weight).

Denoting by $v(r)$ the variance estimator for the rth ($r = 1(1)1{,}000$) sample within a group of 1,000 samples, we calculate

$$\bar{v}_L = \frac{1}{1{,}000}\sum_{r=1}^{1{,}000} v_L(r), \quad \bar{v}_J = \frac{1}{1{,}000}\sum_{r=1}^{1{,}000} v_J(r)$$

$$\bar{v}_{J\pi} = \frac{1}{1{,}000}\sum_{r=1}^{1{,}000} v_{J\pi}(r), \quad \bar{v}_{JX} = \frac{1}{1{,}000}\sum_{r=1}^{1{,}000} v_{JX}(r)$$

separately for all 10 groups of samples and compare results, conditionally for a fixed average value of $(\sum_{s_1} x_i/\pi_{1i})/(\sum_{s_2} x_i/\pi_{1i}\pi_{2i}(s_1))$ to the mean squared error given by

$$\text{MSE} = \frac{1}{10{,}000}\sum_{r=1}^{10{,}000} (t_G(r) - \bar{t}_G)^2$$

where

$$\bar{t}'_G = \frac{1}{10{,}000}\sum_{r=1}^{10{,}000} t'_G(r)$$

Relevant results in this context are given in Table 1.

4.1.1. Remark

From Table 1 and Figure 1 it is evident that, in the example given, the weighted jackknife variance estimator is better than either the linearized or the traditional jackknife variance estimator because the range about the MSE value is smallest for the weighted jackknife variance estimator. Also the performance of the weighted jackknife variance estimator for two different choices of the weights appears to be comparable,

Table 1. Values of four estimated variances over 1,000 samples when N = 284, n′ = 70, n = 20, and MSE = 2,349.46

$\left(\frac{\sum_{s_1} \frac{x_i}{\pi_{1i}}}{\sum_{s_2} \frac{x_i}{\pi_{1i}\pi_{2i}(s_1)}}\right)$	$\bar{v}_L$	$\bar{v}_J$	$\bar{v}_{J\pi}$	$\bar{v}_{Jx}$
0.68	1,364.04	1,702.37	1,962.29	1,964.23
0.78	1,680.12	2,080.31	2,132.60	2,098.69
0.86	1,956.78	2,189.29	2,203.69	2,196.35
0.90	2,102.39	2,255.62	2,301.04	2,295.85
0.96	2,242.21	2,301.03	2,322.46	2,310.29
1.00	2,302.42	2,370.10	2,356.61	2,332.65
1.04	2,511.43	2,468.65	2,398.79	2,443.22
1.12	2,830.41	2,732.44	2,460.56	2,502.04
1.25	3,267.63	2,902.32	2,601.02	2,731.42
1.40	3,890.25	3,109.40	2,756.22	2,895.92

although it may be more important to compensate for extreme sampling weights than to compensate for extreme observations for this set of data. (Results may vary noticeably with other sample sizes and inclusion probabilities and other data. This is yet to be determined. See Section 5)

Next we compare the performance of these variance estimators on the basis of Actual Coverage Percentage (ACP), which gives the percentage of cases in which the confidence interval (CI)

$$t_G \mp \tau_{\alpha/2}\sqrt{v}$$

covers the actual value of the finite population total Y. The closer is the percentage of coverage to $100(1-\alpha)$, the better. To justify the use of the normal score, $\tau_{\alpha/2}$,

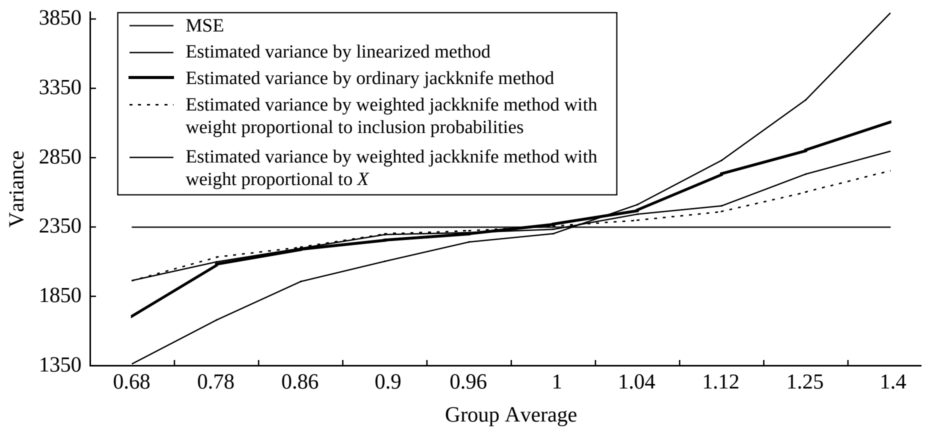

Fig. 1. Comparison of variance estimates by different methods with MSE

Table 2. Values of ACPs with $\alpha = 0.05$ for four variance estimates over 1,000 samples, when $N = 284$, $n' = 70$, $n = 20$, where MSE = 2,349.46 and $(\sqrt{\beta_1}, \beta_2) = (0.015, 2.983)$

$\left(\dfrac{\sum_{s_1} \frac{x_i}{\pi_{1i}}}{\sum_{s_2} \frac{x_i}{\pi_{1i}\pi_{2i}(s_1)}}\right)$	ACPs corresponding to			
	$\bar{v}_L$	$\bar{v}_J$	$\bar{v}_{J\pi}$	$\bar{v}_{Jx}$
0.68	81.2	90.2	92.6	92.6
0.78	87.6	92.1	93.8	93.1
0.86	89.1	93.0	94.3	94.0
0.90	91.2	93.3	94.8	94.3
0.96	93.8	94.4	94.9	94.7
1.00	94.3	95.3	95.1	94.9
1.04	96.8	96.5	95.5	95.7
1.12	98.0	97.8	96.0	96.6
1.25	99.4	98.9	96.6	97.0
1.40	99.8	99.6	97.1	98.0

we calculate the value of

$$\beta_1 = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{t'_G(r) - \bar{t}'_G}{s_G}\right)^3 \quad \text{and} \quad \beta_2 = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{t'_G(r) - \bar{t}'_G}{s_G}\right)^4$$

where $t'_G(r)$ is the value of t'_G based on the rth sample, $\bar{t}'_G = (1/R)\sum_{r=1}^{R} t'_G(r)$ and $s_G = \sqrt{(1/R)\sum_{r=1}^{R} (t'_G(r) - \bar{t}_G)^2}$.

It is noted that for R (= 10,000) samples, the value of $(\sqrt{\beta_1}, \beta_2)$ is (0.015, 2.983), which is not far from (0, 3), indicating normality (Table 2).

4.1.2. Remark

From the above table it appears that with respect to ACPs, the performance of the weighted jackknife variance estimator is very good though the ordinary jackknife variance estimator is a close rival.

4.1.3. Note

A similar simulation experiment was performed for an earlier draft of this article, using substantially fewer replications yet yielded similar results. Therefore our results seem valid for the present set of data.

5. Conclusion and Future Topics

From the results obtained in the limited simulation exercise reported in this article, the weighted jackknife method appears promising indeed. We are applying the weighted jackknife technique to a variety of data with the goal of exploring a solid theoretical foundation that may better explain why this new technique yielded a performance superior to that of either of the other variance estimation techniques. Regarding the choice of weight, a general rule is to be framed so that the weight chosen will compensate to some extent for both the extreme observations and extreme weights simultaneously. These are topics of our research which are expected to be reported later.

Appendix 1: Calculation of $v_J(t'_G)$

$$v_J(t'_G) = \frac{1}{n'(n'-1)}\sum_{i=1}^{n'}(t'_i - t_{JK})^2 = \frac{n'-1}{n'}\sum_{i=1}^{n'}\left\{t_{n'-1}(-j) - \frac{1}{n'}\sum_{j=1}^{n'} t_{n'-1}(-j)\right\}$$

where

$$t_{n'-1}(-j) = \begin{cases} y_R(j) + \dfrac{\sum_{s_2} Q_i(s_1)x_iy_i - Q_j(s_1)x_jy_j}{\sum_{s_2} Q_i(s_1)x_i^2 - Q_j(s_1)x_j^2}(x'_R(j) - x_R(j)) & \text{if } j \in s_2 \\ y_R(j) + B'_Q(s_1)(x'_R(j) - x_R(j)) & \text{if } j \in s_1 - s_2 \end{cases}$$

$$y_R(j) = \begin{cases} \sum_{s_2} \dfrac{y_i}{\pi_{1i}\pi_{2i}(s_1)} - \dfrac{y_j}{\pi_{1j}\pi_{2j}(s_1)} & \text{if } j \in s_2 \\ \sum_{s_2} \dfrac{y_i}{\pi_{1i}\pi_{2i}(s_1)} & \text{if } j \in s_1 - s_2 \end{cases}$$

$$x_R(j) = \begin{cases} \sum_{s_2} \dfrac{x_i}{\pi_{1i}\pi_{2i}(s_1)} - \dfrac{x_j}{\pi_{1j}\pi_{2j}(s_1)} & \text{if } j \in s_2 \\ \sum_{s_2} \dfrac{x_i}{\pi_{1i}\pi_{2i}(s_1)} & \text{if } j \in s_1 - s_2 \end{cases}$$

and

$$x'_R(j) = \sum_{s_1} \frac{x_i}{\pi_{1i}} - \frac{x_j}{\pi_{1j}} \quad \text{if } j \in s_1$$

If $j \in s_2$

$$t_{n'-1}(-j) = \left(\sum_{s_2} \frac{y_i}{\pi_{1i}\pi_{2i}(s_1)} - \frac{y_j}{\pi_{1j}\pi_{2j}(s_1)}\right) + \left(\frac{\sum_{s_2} Q_i(s_1)x_iy_i - Q_j(s_1)x_jy_j}{\sum_{s_2} Q_i(s_1)x_i^2 - Q_j(s_1)x_j^2}\right)$$
$$\times\left\{\left(\sum_{s_1}\frac{x_i}{\pi_{1i}} - \frac{x_j}{\pi_{1j}}\right) - \left(\sum_{s_2}\frac{x_i}{\pi_{1i}\pi_{2i}(s_1)} - \frac{x_j}{\pi_{1j}\pi_{2j}(s_1)}\right)\right\}$$

Assuming

$$\left|\frac{Q_j(s_1)x_jy_j}{\sum_{s_2} Q_i(s_1)x_iy_i}\right| < 1 \quad \text{and} \quad \left|\frac{Q_j(s_1)x_j^2}{\sum_{s_2} Q_i(s_1)x_i^2}\right| < 1$$

we get

$$t_{n'-1}(-j) = t_{n'} - g_{sj}\frac{y_j - B'_Q(s_1)x_j}{\pi_{1j}\pi_{2j}(s_1)} - B'_Q(s_1)\frac{x_j}{\pi_{1j}}$$

where

$$g_{si} = 1 + \frac{Q_i(s_1)x_i\pi_{1i}\pi_{2i}(s_1)}{\sum_{s_1} Q_i(s_1)x_i^2}\left(\sum_{s_1}\frac{x_i}{\pi_{1i}} - \sum_{s_2}\frac{x_i}{\pi_{1i}\pi_{2i}(s_1)}\right)$$

and

$$\frac{1}{n'}\sum_{s_2} t_{n'-1}(-j) = \frac{1}{n'}\left[(nt_{n'}) - \sum_{s_2} g_{si}\frac{y_i - B'_Q(s_1)x_i}{\pi_{1i}\pi_{2i}(s_1)} - B'_Q(s_1)\sum_{s_2}\frac{x_j}{\pi_{1j}}\right]$$

$$= \frac{n}{n'}t_{n'} - \frac{1}{n'}\sum_{s_2}\left\{1 + \frac{Q_i(s_1)x_i\pi_{1i}\pi_{2i}(s_1)}{\sum_{s_1} Q_i(s_1)x_i^2}\left(\sum_{s_1}\frac{x_i}{\pi_{1i}} - \sum_{s_2}\frac{x_i}{\pi_{1i}\pi_{2i}(s_1)}\right)\right\}\frac{y_i - B'_Q(s_1)x_i}{\pi_{1i}\pi_{2i}(s_1)}$$

$$-\frac{1}{n'}B'_Q(s_1)\sum_{s_2}\frac{x_j}{\pi_{1j}}$$

$$= \frac{n}{n'}t_{n'} - \frac{1}{n'}\sum_{s_2}\frac{y_i - B'_Q(s_1)x_i}{\pi_{1i}\pi_{2i}(s_1)} + \left(\sum_{s_1}\frac{x_i}{\pi_{1i}} - \sum_{s_2}\frac{x_i}{\pi_{1i}\pi_{2i}(s_1)}\right)\frac{\sum_{s_2} Q_i(s_1)x_i(y_i - B'_Q(s_1)x_i)}{\sum_{s_2} Q_i(s_1)x_i^2}$$

$$-\frac{1}{n'}B'_Q(s_1)\sum_{s_2}\frac{x_j}{\pi_{1j}}$$

$$= \frac{n}{n'}t_{n'} - \frac{1}{n'}\sum_{s_2}\frac{y_i - B'_Q(s_1)x_i}{\pi_{1i}\pi_{2i}(s_1)} + 0 - \frac{1}{n'}B_Q(s_1)\sum_{s_2}\frac{x_j}{\pi_{1j}}$$

$$= \frac{n}{n'}t_{n'} - \frac{1}{n'}t_{n'} + \frac{1}{n'}B'_Q(s_1)\sum_{s_1}\frac{x_i}{\pi_{1i}} - \frac{1}{n'}B'_Q(s_1)\sum_{s_2}\frac{x_j}{\pi_{1j}}$$

If $j \in s_1 - s_2$

$$t_{n'-1}(-j) = \sum_{s_2}\frac{y_i}{\pi_{1i}\pi_{2i}(s_1)} + B'_Q(s_1)\left[\left\{\sum_{s_1}\frac{x_i}{\pi_{1i}} - \frac{x_i}{\pi_{1i}}\right\} - \sum_{s_2}\frac{x_i}{\pi_{1i}\pi_{2i}(s_1)}\right] = t_{n'} - B'_Q(s_1)\frac{x_j}{\pi_{1j}}$$

and

$$\frac{1}{n'}\sum_{s_1-s_2} t_{n'-1}(-j) = \frac{n'-n}{n'}t_{n'} - B'_Q(s_1)\sum_{s_1-s_2}\frac{x_i}{\pi_{1i}}$$

Therefore

$$v_J(t'_G) = \frac{n'-1}{n'}\sum_{i=1}^{n'}\left\{t_{n'-1}(-j) - \frac{1}{n'}\sum_{j=1}^{n'} t_{n'-1}(-j)\right\}^2$$

$$= \frac{n'-1}{n'}\left[\sum_{s_1}\left\{-g_{si}\frac{y_i - B'_Q(s_1)x_i}{\pi_{1i}\pi_{2i}(s_1)} - B'_Q(s_1)\frac{x_i}{\pi_{1i}} + B'_Q(s_1)\frac{1}{n'}\sum_{s_1}\frac{x_i}{\pi_{1i}}\right\}^2\right]$$

So for large n'

$$v_J(t'_G)=\sum_{s_2}\left[g_{si}\frac{(y_i-B'_Q(s_1)x_i)}{\pi_{1i}\pi_{2i}(s_1)}\right]^2+B'_Q(s_1)\sum_{s_1}\left(\frac{x_i}{\pi_{1i}}-\frac{1}{n'}\sum_{s_1}\frac{x_i}{\pi_{1i}}\right)^2$$
$$+2B'_Q(s_1)\sum_{s_2}g_{si}\frac{(y_i-B'_Q(s_1)x_i)}{\pi_{1i}\pi_{2i}(s_1)}\left(\frac{x_i}{\pi_{1i}}-\frac{1}{n'}\sum_{s_1}\frac{x_i}{\pi_{1i}}\right)$$

Appendix 2: Calculation of $v_L(t'_R)$ and $v_J(t'_R)$

We already obtained the expression for $v_L(t'_G)$ as

$$v_L(t'_G)=\sum\sum_{s_2}\frac{\Delta_{1ij}}{\pi_{1ij}\pi_{2ij}(s_1)}\left(\frac{e_i(s_1)}{\pi_{1i}}-\frac{e_j(s_1)}{\pi_{1j}}\right)^2+B'_Q2(s_1)\sum\sum_{s_1}\frac{\Delta_{1ij}}{\pi_{1ij}}\left(\frac{x_i}{\pi_{1i}}-\frac{x_j}{\pi_{1j}}\right)^2$$
$$+2B'_Q(s_1)\sum\sum_{s_2}\frac{\Delta_{1ij}}{\pi_{1ij}\pi_{2ij}(s_1)}\left(\frac{e_i(s_1)}{\pi_{1i}}-\frac{e_j(s_1)}{\pi_{1j}}\right)\left(\frac{x_i}{\pi_{1i}}-\frac{x_j}{\pi_{1j}}\right)$$
$$+\sum\sum_{s_2}\frac{\Delta_{2ij}(s_1)}{\pi_{2ij}(s_1)}\left(\frac{e_i(s_1)}{\pi_{1i}\pi_{2i}(s_1)}-\frac{e_j(s_1)}{\pi_{1j}\pi_{2j}(s_1)}\right)^2$$

With the choice of $Q_i(s_1)=1/(x_i\pi_{1i}\pi_{2i}(s_1))$ we get

$$B'_Q(s_1)=\frac{\sum_{s_2}\frac{y_i}{\pi_{1i}\pi_{2i}(s_1)}}{\sum_{s_2}\frac{x_i}{\pi_{1i}\pi_{2i}(s_1)}}=B'_R(s_1)\ \text{(say)}\quad\text{and}\quad e_i(s_1)=y_i-B'_R(s_1)$$

So the above variance estimator can be written as

$$v_L(t'_R)=\sum\sum_{s_2}\frac{\Delta_{1ij}}{\pi_{1ij}\pi_{2ij}(s_1)}\left(\frac{y_i-B'_R(s_1)x_i}{\pi_{1i}}-\frac{y_j-B'_R(s_1)x_j}{\pi_{1j}}\right)^2$$
$$+B_R'^2(s_1)\sum\sum_{s_1}\frac{\Delta_{1ij}}{\pi_{1ij}}\left(\frac{x_i}{\pi_{1i}}-\frac{x_j}{\pi_{1j}}\right)^2$$
$$+2B_R'^2(s_1)\sum\sum_{s_2}\frac{\Delta_{1ij}}{\pi_{1ij}\pi_{2ij}(s_1)}\left(\frac{y_i-B'_R(s_1)x_i}{\pi_{1i}}-\frac{y_j-B'_R(s_1)x_j}{\pi_{1j}}\right)\left(\frac{x_i}{\pi_{1i}}-\frac{x_j}{\pi_{1j}}\right)$$
$$+\sum\sum_{s_2}\frac{\Delta_{2ij}(s_1)}{\pi_{2ij}(s_1)}\left(\frac{y_i-B'_R(s_1)x_i}{\pi_{1i}\pi_{2i}(s_1)}-\frac{y_j-B'_R(s_1)x_j}{\pi_{1j}\pi_{2j}(s_1)}\right)^2$$
$$=\sum\sum_{s_2}\frac{\Delta_{1ij}}{\pi_{1ij}\pi_{2ij}(s_1)}\left(\frac{y_i-B'_R(s_1)x_i}{\pi_{1i}}-\frac{y_j-B'_R(s_1)x_j}{\pi_{1j}}\right)^2$$
$$+\sum\sum_{s_2}\frac{\Delta_{2ij}(s_1)}{\pi_{2ij}(s_1)}\left(\frac{y_i-B'_R(s_1)x_i}{\pi_{1i}\pi_{2i}(s_1)}-\frac{y_i-B'_R(s_1)x_j}{\pi_{1j}\pi_{2j}(s_1)}\right)^2$$
$$+B_R'^2(s_1)\sum\sum_{s_1}\frac{\Delta_{1ij}}{\pi_{1ij}}\left(\frac{x_i}{\pi_{1i}}-\frac{x_j}{\pi_{1j}}\right)^2$$
$$+2B'_R(s_1)\sum\sum_{s_2}\frac{\Delta_{1ij}}{\pi_{1ij}\pi_{2ij}(s_1)}\left(\frac{y_i-B'_R(s_1)x_i}{\pi_{1i}}-\frac{y_j-B'_R(s_1)x_j}{\pi_{1j}}\right)\left(\frac{x_i}{\pi_{1i}}-\frac{x_j}{\pi_{1j}}\right)$$

Under equal probability sampling, we take

$$\pi_{1i} = \frac{n'}{N}, \quad \pi_{2i}(s_1) = \frac{n}{n'}, \quad \pi_{1ij} = \frac{n'(n'-1)}{N(N-1)} \quad \text{and} \quad \pi_{2ij}(s_1) = \frac{n(n-1)}{n'(n'-1)}$$

So we may write

$$\begin{aligned} \Delta_{1ij} &= \pi_{1i}\pi_{1j} - \pi_{1ij} = \frac{n'}{N} \cdot \frac{(N-n')}{N(N-1)}, \\ \Delta_{2ij}(s_1) &= \pi_{2i}(s_1)\pi_{2j}(s_1) - \pi_{2ij}(s_1) = \frac{n}{n'} \cdot \frac{(n'-n)}{n'(n'-1)}, \\ B'_R(s_1) &= \frac{\bar{y}}{\bar{x}} = \hat{R} \quad \text{and} \quad e_i(s_1) = y_i - \hat{R}x_i \end{aligned}$$

Using the above results we get

$$\begin{aligned} v_L(t'_R) &= N^2 \cdot \frac{N-n'}{Nn'} \frac{1}{n(n-1)} \sum\sum_{s_2} \left\{(y_i - \hat{R}x_i) - (y_j - \hat{R}x_j)\right\}^2 \\ &\quad + N^2 \cdot \frac{n'-n}{n'n} \frac{1}{n(n-1)} \sum\sum_{s_2} \left\{(y_i - \hat{R}x_i) - (y_j - \hat{R}x_j)\right\}^2 \\ &\quad + \hat{R}^2 N^2 \cdot \frac{N-n'}{Nn'} \frac{1}{n'(n'-1)} \sum\sum_{s_1} (x_i - x_j)^2 \\ &\quad + 2\hat{R}N^2 \cdot \frac{N-n'}{Nn'} \frac{1}{n(n-1)} \sum\sum_{s_2} \left\{(y_i - \hat{R}x_i) - (y_j - \hat{R}x_j)\right\}(x_i - x_j) \\ &= N^2 \left[\left(\frac{N-n}{Nn}\right) \frac{1}{(n-1)} \sum_{s_2} (y_i - \hat{R}x_i)^2 + \hat{R}^2 \left(\frac{N-n'}{Nn'}\right) \frac{1}{n'-1} \sum_{s_1} (x_i - \bar{x})^2 \right. \\ &\quad \left. + 2\hat{R}\left(\frac{N-n'}{Nn'}\right) \frac{1}{(n-1)} \sum_{s_2} (y_i - \hat{R}x_i)(x_i - \bar{x}) \right] \end{aligned}$$

Now ignoring the finite population correction, the variance estimator of t'_R takes the form

$$v_L(t'_R) \cong N^2 \left[\frac{s_e^2}{n} + \hat{R}^2 \frac{{s'}_x^2}{n'} + 2\hat{R} \frac{s_{ex}}{n'} \right]$$

Now let us refer to the expression for $v_J(t'_G)$ obtained in Appendix 1.

With the choice $Q_i(s_1) = 1/(x_i\pi_{1i}\pi_{2i}(s_1))$ we get

$$\begin{aligned} g_{si} &= 1 + \frac{1}{\sum_{s_2} \frac{x_i}{\pi_{1i}\pi_{2i}(s_1)}} \left(\sum_{s_1} \frac{x_i}{\pi_{1i}} - \sum_{s_2} \frac{x_i}{\pi_{1i}\pi_{2i}(s_1)} \right) \\ &= \frac{\sum_{s_1} \frac{x_i}{\pi_{1i}}}{\sum_{s_2} \frac{x_i}{\pi_{1i}\pi_{2i}(s_1)}} = g_{si}(R) \quad \text{say} \end{aligned}$$

and we get

$$v_J(t'_R) = \sum_{s_2}\left[g_{si}(R)\frac{(y_i - B'_R(s_1)x_i)}{\pi_{1i}\pi_{2i}(s_1)}\right]^2 + B'^2_R(s_1)\sum_{s_1}\left(\frac{x_i}{\pi_{1i}} - \frac{1}{n'}\sum_{s_1}\frac{x_i}{\pi_{1i}}\right)^2$$
$$+ 2B'_R(s_1)\sum_{s_2} g_{si}(R)\frac{(y_i - B'_R(s_1)x_i)}{\pi_{1i}\pi_{2i}(s_1)}\left(\frac{x_i}{\pi_{1i}} - \frac{1}{n'}\sum_{s_1}\frac{x_i}{\pi_{1i}}\right)$$

Under equal probability sampling, $g_{si}(R)$ reduces to $(\bar{x}'/\bar{x})$, and the jackknife variance estimator takes the following form:

$$v_J(t'_R) \cong N^2\left[\left(\frac{\bar{x}'}{\bar{x}}\right)^2\frac{s_e^2}{n} + \hat{R}^2\frac{s'^2_x}{n'} + 2\hat{R}\left(\frac{\bar{x}'}{\bar{x}}\right)\frac{s_{ex}}{n'}\right]$$

6. References

Basu, D. (1971). An Essay on the Logical Foundations of Survey Sampling. In Foundations of Statistical Inference. Godambe, V.P. and Sprott, D.A. (eds). Holt, Rinehart, Winston, Toronto. Part I, 203–242.

Brewer, K.R.W. (1979). A Class of Robust Sampling Designs for Large Scale Surveys. Journal of the American Statistical Association, 74, 911–915.

Deville, J.C. and Särndal, C.E. (1992). Calibration Estimators in Survey Sampling. Journal of the American Statistical Association, 87, 376–382.

Durbin, J. (1959). A Note on the Application of Quenouille's Method of Bias Reduction to the Estimation of Ratios. Biometrika, 46, 477–480.

Hajék, J. (1971). Comment on a Paper by Basu, D. In Foundations of Statistical Inference, Holt, Reinhart, Winston, Toronto. 203–242.

Midzuno, H. (1952). On the System with Probabilities Proportionate to Sum of Sizes. Annals of the Institute of Statistical Mathematics, 3, 99–107.

Quenouille, M.H. (1956). Notes on Bias in Estimation. Biometrika, 43, 353–360.

Rao, J.N.K. (1965). A Note on Estimation of Ratios by Quenouille's Method. Biometrika, 52, 647–649.

Rao, J.N.K. and Sitter, R.R. (1995). Variance Estimator Under Two Phase Sampling with Application to Imputation for Missing Data. Biometrika, 82(2), 453–460.

Rao, J.N.K. and Webster, J.T. (1966). On Two Methods of Bias Reduction in the Estimation of Ratios. Biometrika, 53, 571–577.

Roy, D. and Safiquzzaman, M. (2000). Jackknifing a General Class of Estimators. Brazilian Journal of Probability and Statistics, 14, 47–60.

Roy, D. and Safiquzzaman, M. (2002). Variance Estimation by Jackknifing Method: A New Approach with Reference to Varying Probability Sampling. Calcutta Statistical Association Bulletin, Vol. 52 (Special 4th Triennial Proceedings Volume) Nos 205–208, 335–348.

Roy, D. and Safiquzzaman, M. (2003). Jackknifing a General Class of Estimators and Variance Estimation Under Two Phase Sampling. Metron, LXI, 53–74.

Särndal, C.E. (1980). On π-inverse Weighting Versus Best Linear Weighting in Probability Sampling. Biometrika, 67, 639–650.

Särndal, C.E., Swensson, B., and Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer-Verlag.

Särndal, C.E. and Wright, R.L. (1984). Cosmetic Form of Estimators in Survey Sampling. Scandinavian Journal of Statistics, 11, 146–156.

Tukey, J. (1958). Bias and Confidence in Not Quite Large Sample. Annuals of Mathematical Statistics, 29, 614.

Wright, R.L. (1983). Finite Population Sampling with Multivariate Auxiliary Information. Journal of American Statistical Association, 78, 879–884.

Received July 2003
Revised September 2005