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

Article

Some contributions to jackknifing two-phase sampling estimators

by Patrick J. Farrell and Sarjinder Singh

June 2010

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2010
Vol. 36, No. 1, pp. 57-68
Statistics Canada, Catalogue No. 12-001-X

Some contributions to jackknifing two-phase sampling estimators

Patrick J. Farrell and Sarjinder Singh [1]

Abstract

In this paper, the problem of estimating the variance of various estimators of the population mean in two-phase sampling has been considered by jackknifing the two-phase calibrated weights of Hidiroglou and Särndal (1995, 1998). Several estimators of population mean available in the literature are shown to be the special cases of the technique developed here, including those suggested by Rao and Sitter (1995) and Sitter (1997). By following Raj (1965) and Srivenkataramana and Tracy (1989), some new estimators of the population mean are introduced and their variances are estimated through the proposed jackknife procedure. The variance of the chain ratio and regression type estimators due to Chand (1975) are also estimated using the jackknife. A simulation study is conducted to assess the efficiency of the proposed jackknife estimators relative to the usual estimators of variance.

Key Words: Auxiliary information; Calibration; Estimation of mean and variance; Jackknife; Two-phase sampling.

1. Introduction

Hidiroglou and Särndal (1995, 1998) have pointed out that two-phase sampling for the estimation of finite population attributes is a powerful and cost-effective technique, and hence plays an eminent role in survey sampling. Two-phase sampling can be described as follows. Consider a finite population that we shall denote by $\Omega = \{1, 2, \ldots, i, \ldots, N\}$. Suppose that information is available on a variable Z across the entire population; that is, the values Z_i for all $i = 1, \ldots, N$, are known, implying that the population mean, $\bar{Z}$, is also known. A first-phase probability sample s_1, $s_1 \subset \Omega$, of size m is drawn from the population with selection probabilities π_{1i}. Thus, the first-phase sampling weights can be defined as $d_{1i} = 1/\pi_{1i}$. Assume that for this sample, information is collected on a variable X, which is then paired with the information on Z for each of the m units, giving rise to the data $\{(x_i, z_i) \mid i \in s_1\}$ for $i = 1, \ldots, m$. Once the first-phase sample s_1 has been drawn, a second-phase sample s_2, $s_2 \subset s_1 \subset \Omega$, of size n is selected from s_1 with selection probabilities $\pi_{2i} = \pi_{i|s_1}$, allowing for the second-phase sampling weights to be defined as $d_{2i} = 1/\pi_{2i}$. In the second-phase sample, information is now collected on a variable Y for each selected unit. This information is linked to that previously available on Z and X for these units, giving rise to the data $\{(x_i, y_i, z_i) \mid i \in s_2\}$ for $i = 1, \ldots, n$. Suppose that interest lies in estimating the population mean $\bar{Y}$, and on the variance of the estimator employed.

Let $w_{1i}^o = d_{1i} / \sum_{i \in s_1} d_{1i}$ denote the first-phase normalized original design weights. The usual estimator of the population mean $\bar{X}$ is given by

$$\hat{\bar{X}}_1^o = \sum_{i \in s_1} w_{1i}^o x_i,$$

while a calibrated first-phase estimator of $\bar{X}$ is

$$\hat{\bar{X}}_1^c = \sum_{i \in s_1} w_{1i}^c x_i,$$

where the w_{1i}^c are calibrated weights such that the chi-square distance function

$$D_1 = \sum_{i \in s_1} \{(w_{1i}^c - w_{1i}^o)^2 / (w_{1i}^o q_{1i})\}, \qquad (1.1)$$

is minimized subject to

$$\sum_{i \in s_1} w_{1i}^c z_i = \bar{Z}. \qquad (1.2)$$

In (1.1), the q_{1i} are a set of suitably chosen weights. Minimization of (1.1) subject to (1.2) leads to the first-phase calibrated weights

$$w_{1i}^c = w_{1i}^o + \left\{(q_{1i} w_{1i}^o z_i) \Big/ \left(\sum_{i \in s_1} q_{1i} w_{1i}^o z_i^2\right)\right\}\left(\bar{Z} - \sum_{i \in s_1} w_{1i}^o z_i\right).$$

Thus, a first-phase calibrated estimator of $\bar{X}$ is given by

$$\hat{\bar{X}}_1^c = \sum_{i \in s_1} w_{1i}^o x_i + \hat{\beta}_1 \left(\bar{Z} - \sum_{i \in s_1} w_{1i}^o z_i\right),$$

where

$$\hat{\beta}_1 = \left(\sum_{i \in s_1} q_{1i} w_{1i}^o x_i z_i\right) \Big/ \left(\sum_{i \in s_1} q_{1i} w_{1i}^o z_i^2\right).$$

Now, let $w_{2i}^o = d_{1i} d_{2i} / \sum_{i \in s_2} d_{1i} d_{2i}$ denote the second-phase normalized design weights. The usual estimator of $\bar{Y}$ is given by

1. Patrick J. Farrell, School of Mathematics and Statistics, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, Canada, K1S 5B6. E-mail: pfarrell@math.carleton.ca; Sarjinder Singh, Department of Mathematics, Texas A&M University - Kingsville, Kingsville, Texas, U.S.A., 78363. E-mail: sarjinder@yahoo.com.

$$\hat{\bar{Y}}_2^o = \sum_{i \in s_2} w_{2i}^o y_i.$$

Let us consider the second-phase calibrated estimator of $\bar{Y}$ as

$$\hat{\bar{Y}}^c = \sum_{i \in s_2} w_{2i}^c y_i, \quad (1.3)$$

where the w_{2i}^c are the second-phase calibrated weights such that the chi-square distance function

$$D_2 = \sum_{i \in s_2} \{(w_{2i}^c - w_{2i}^o)^2 / (w_{2i}^o q_{2i})\}, \quad (1.4)$$

is minimized subject to the calibration constraint

$$\sum_{i \in s_2} w_{2i}^c x_i = \hat{\bar{X}}_1^c. \quad (1.5)$$

Minimization of (1.4) subject to (1.5) leads to the second-phase calibrated weights

$$w_{2i}^c = w_{2i}^o + \left\{(q_{2i} w_{2i}^o x_i) \Big/ \left(\sum_{i \in s_2} q_{2i} w_{2i}^o x_i^2\right)\right\} \left(\hat{\bar{X}}_1^c - \sum_{i \in s_2} w_{2i}^o x_i\right).$$

Thus, the second-phase calibrated estimator of $\bar{Y}$ specified in (1.3) can be written as

$$\hat{\bar{Y}}^c = \hat{\bar{Y}}_2^o + \hat{\beta}_2(\hat{\bar{X}}_1^o - \hat{\bar{X}}_2^o) + \hat{\beta}_1 \hat{\beta}_2(\bar{Z} - \hat{\bar{Z}}_1^o), \quad (1.6)$$

where $\hat{\bar{Z}}_1^o = \sum_{i \in s_1} w_{1i}^o z_i$, $\hat{\bar{X}}_1^o = \sum_{i \in s_1} w_{1i}^o x_i$, $\hat{\bar{X}}_2^o = \sum_{i \in s_2} w_{2i}^o x_i$, $\hat{\bar{Y}}_2^o = \sum_{i \in s_2} w_{2i}^o y_i$, and

$$\hat{\beta}_2 = \left(\sum_{i \in s_2} q_{2i} w_{2i}^o x_i y_i\right) \Big/ \left(\sum_{i \in s_2} q_{2i} w_{2i}^o x_i^2\right).$$

Hidiroglou and Särndal (1995, 1998) and Singh (2000) have considered the problem of estimating the variance of the calibrated estimator $\hat{\bar{Y}}^c$ in (1.6) by using a design-based approach. In a more general context, Rao and Sitter (1995) and Sitter (1997) have pointed out that under simple random sampling without replacement (SRSWOR), a jackknife technique can be used to estimate the variances of the ratio and regression estimators for a population mean. These authors have also reported that the use of the jackknife for estimating variance is more convenient and efficient than the traditional techniques based on estimates of moments.

Of late, a number of authors have investigated the use of jackknife procedures for estimating variances (See Arnab and Singh 2006, Berger 2007, Berger and Skinner 2005, Chen and Shao 2001, and Kovar and Chen 1994). Fuller (1998), Kim, Navarro and Fuller (2000, 2006), Kim and Sitter (2003), and Kott and Stukel (1997) have suggested an approach for estimating the variance in two-stage sampling. Fuller (1998) and Kim and Sitter (2003) address the regression estimator. In particular, consider the generalized regression estimator of population total

$$\hat{Y}_{\mathrm{DS}} = \sum_{i \in s_2} \alpha_i y_i,$$

due to Deville and Särndal (1992). Following Kim *et al.* (2000, 2006), for each $k \in s_2$, specify the jackknife estimator of population total as

$$\hat{Y}_{\mathrm{Kim}} = \sum_{i \in s_2 \setminus k} \alpha_i^{(k)} y_i, \quad (1.7)$$

and the chi-square distance between the design and calibration weights as

$$D_{(k)} = (1/2) \sum_{i \in s_2 \setminus k} \{(\alpha_i^{(k)} - w_i^{(k)} w_i^{*(k)})^2 / (w_i^{(k)} q_i^{(k)})\}. \quad (1.8)$$

Minimizing (1.8) subject to the condition

$$\sum_{i \in s_2 \setminus k} \alpha_i^{(k)} x_i = \sum_{i \in s_1 \setminus k} w_i^{(k)} x_i,$$

leads to jackknifed calibrated weights given by

$$\alpha_i^{(k)} = w_i^{(k)} w_i^{*(k)} + \left\{(w_i^{(k)} q_i^{(k)} x_i) \Big/ \left(\sum_{i \in s_2 \setminus k} w_i^{(k)} q_i^{(k)}\right)\right\} \left\{\sum_{i \in s_2 \setminus k} w_i^{(k)} x_i - \sum_{i \in s_2 \setminus k} w_i^{(k)} w_i^{*(k)} x_i\right\}.$$

It would appear that Kim *et al.* (2006) readjusted these weights as

$$\alpha_i^{(k)} = \begin{cases} \alpha_i^{(k)} & \text{if } k \in s_2 \\ w_i^{(k)} & \text{if } j \in (s_1 - s_2). \end{cases}$$

For such a readjustment, the estimator in (1.7) is equivalent to that of Rao and Sitter (1995).

In the present paper, we consider a new jackknife technique to estimate the variance of the estimator $\hat{\bar{Y}}^c$ under the two-phase setup by following Hidiroglou and Särndal (1995, 1998). Similar to Kim *et al.* (2006), the estimator proposed by Rao and Sitter (1995) is shown to be a special case of the proposed method. However, our approach differs from that of Fuller (1998) Kim and Sitter (2003), Kim *et al.* (2000, 2006) in that we consider calibration at both the first and second phases, thus allowing for the development of the technique for chain ratio and chain regression type estimators. We also investigate, via a simulation study, the efficiency of the jackknife estimators of variance relative to the usual estimators.

2. Estimation of variance using jackknifing

In what follows, we assume that a single stage design is employed at both of the two phases in the sampling process.

Let $\hat{\bar{Y}}^c(j)$ be a calibrated estimator of the population mean, $\bar{Y}$, obtained by dropping the j^{th} unit from the sample s_1 of m units. We prove in the Appendix that the jackknife estimator of the population mean in two phase-sampling can be written as

$$\hat{\bar{Y}}^c(j) = \begin{cases} \hat{\bar{Y}}_2^o(j) + \hat{\beta}_2(j)\,\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_2^o(j)\} \\ \quad + \hat{\beta}_1(j)\,\hat{\beta}_2(j)\{\bar{Z} - \hat{\bar{Z}}_1^o(j)\} & \text{if } j \in s_2 \\ \\ \hat{\bar{Y}}_2^o + \hat{\beta}_2\,\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_1^o\} \\ \quad + \hat{\beta}_1(j)\,\hat{\beta}_2\{\bar{Z} - \hat{\bar{Z}}_1^o(j)\} & \text{if } j \in (s_1 - s_2) \end{cases} \tag{2.1}$$

where the quantity $\hat{\bar{Z}}_1^o(j) = \hat{\bar{Z}}_1^o + \{w_{1j}^o/(1 - w_{1j}^o)\}\{\hat{\bar{Z}}_1^o - z_j\}$, the terms $\hat{\bar{X}}_1^o(j)$, $\hat{\bar{X}}_2^o(j)$, and $\hat{\bar{Y}}_2^o(j)$ are defined in an analogous manner, $\hat{\beta}_1(j) = \hat{\beta}_1 + \{q_{1j} w_{1j}^o z_j (x_j - \hat{\beta}_1 z_j)\}/\{q_{1j} w_{1j}^o z_j^2 - \sum_{i \in s_1} q_{1i} w_{1i}^o z_i^2\}$, and $\hat{\beta}_2(j) = \hat{\beta}_2 + \{q_{2j} w_{2j}^o x_j (y_j - \hat{\beta}_2 x_j)\}/\{q_{2j} w_{2j}^o x_j^2 - \sum_{i \in s_1} q_{2i} w_{2i}^o x_i^2\}$. The modified jackknife estimator of variance of $\hat{\bar{Y}}^c$ is then given by

$$\hat{V}_{\text{JACK}}(\hat{\bar{Y}}^c) = \{(m-1)/m\} \sum_{j \in s_1} \{\hat{\bar{Y}}^c(j) - \hat{\bar{Y}}^c\}^2. \tag{2.2}$$

We show in the appendix that this estimator is consistent.

Note that we can write that

$$\hat{\bar{Y}}^c(j) - \hat{\bar{Y}}^c = \begin{cases} \varepsilon_2(j) + \hat{\beta}_2 \varepsilon_1(j) + \hat{\beta}_2(j)\, d_2(j) \\ \quad + \hat{\beta}_2 \delta_2(j) & \text{if } j \in s_2 \\ \\ \hat{\beta}_2 \varepsilon_1(j) & \text{if } j \in (s_1 - s_2) \end{cases} \tag{2.3}$$

where the terms in (2.3) are given by $\varepsilon_1(j) = \{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_1^o\} - \hat{\beta}_1(j)\{\hat{\bar{Z}}_1^o(j) - \bar{Z}\}$, $\varepsilon_2(j) = \{\hat{\bar{Y}}_2^o(j) - \hat{\bar{Y}}_2^o\} - \hat{\beta}_2(j)\{\hat{\bar{X}}_2^o(j) - \hat{\bar{X}}_2^o\} - \hat{\beta}_1(j)\,\hat{\beta}_2(j)\{\hat{\bar{Z}}_1^o(j) - \bar{Z}\}$, $d_2(j) = \{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_2^o(j)\}$ and $\delta_2(j) = \{\hat{\bar{X}}_2^o(j) - \hat{\bar{X}}_1^o\} - \hat{\beta}_1(j)\{\bar{Z} - \hat{\bar{Z}}_1^o(j)\} - \hat{\beta}_1\{\bar{Z} - \hat{\bar{Z}}_1^o\}$. The $\varepsilon_1(j)$ term is analogous to the error term associated with the regression of the auxiliary variable x_i on z_i, for $i \in s_1$, while $\varepsilon_2(j)$ is analogous to the error term associated with the regression of the study variable y_i on both x_i and z_i simultaneously, for $i \in s_2$. Provided that $j \in s_2$, the $d_2(j)$ term reflects the difference in the jackknife first and second phase sample means for the variable X, while $\delta_2(j)$ denotes an adjustment to $d_2(j)$ obtained by using information on the auxiliary variable Z.

Using (2.3) in (2.2), the jackknife estimator of variance of the estimator $\hat{\bar{Y}}^c$ is given by

$$\begin{aligned} \hat{V}_{\text{JACK}}(\hat{\bar{Y}}^c) = \{(m-1)/m\} \Bigg[& \sum_{j \in s_2} \varepsilon_2^2(j) + \sum_{j \in s_2} \hat{\beta}_2^2(j) d_2^2(j) \\ & + \hat{\beta}_2^2 \sum_{j \in s_2} \delta_2(j)\{\delta_2(j) + 2\varepsilon_1(j)\} \\ & + 2\hat{\beta}_2 \sum_{j \in s_2} \varepsilon_1(j)\varepsilon_2(j) \\ & + 2\hat{\beta}_2 \sum_{j \in s_2} \hat{\beta}_2(j)\, d_2(j)\{\varepsilon_1(j) + \delta_2(j)\} \\ & + \hat{\beta}_2^2 \sum_{j \in s_1} \varepsilon_1^2(j) \Bigg]. \end{aligned} \tag{2.4}$$

Note that the expression given in (2.4) is exact. It can be used to estimate the variance of several estimators available in the literature.

3. Special cases

In the next section, we demonstrate that the estimator proposed by Rao and Sitter (1995), Sitter (1997), Raj (1965), Srivenkataramana and Tracy (1989), Chand (1975), and Ahmed (1997) can be viewed as special cases of the proposed technique.

Case 3.1: Rao and Sitter (1995)

If $\hat{\bar{X}}_1^c = \hat{\bar{X}}_1^o$ (no first-phase calibration is made) and $q_{2i} = 1/x_i$, then the calibrated estimator of $\bar{Y}$ becomes

$$\hat{\bar{Y}}_r^c = \left(\sum_{i \in s_2} w_{2i}^o y_i\right) \left\{ \left(\sum_{i \in s_2} w_{1i}^o x_i\right) \Big/ \left(\sum_{i \in s_2} w_{2i}^o x_i\right) \right\}.$$

If the first-phase sample s_1 is selected according to SRSWOR such that the first-phase design weights are given by $d_{1i} = N/m$, and the second-phase sample s_2 is selected from s_1 by SRSWOR such that $d_{2i} = m/n$, then the calibrated estimator of the population mean becomes

$$\hat{\bar{Y}}_{\text{RS}}^c = \bar{y}(\bar{x}'/\bar{x}), \tag{3.1}$$

where $\bar{y} = \sum_{i \in s_2} y_i/n$, $\bar{x} = \sum_{i \in s_2} x_i/n$, and $\bar{x}' = \sum_{i \in s_1} x_i/m$. The jackknife mechanism in (2.1) becomes

$$\hat{\bar{Y}}_{\text{RS}}^c(j) = \begin{cases} \dfrac{(n\bar{y} - y_j)(m\bar{x}' - x_j)}{(n\bar{x} - x_j)(m-1)} & \text{if } j \in s_2 \\ \\ (\bar{y}/\bar{x})\dfrac{(m\bar{x}' - x_j)}{(m-1)} & \text{if } j \in (s_1 - s_2). \end{cases} \tag{3.2}$$

Setting $\hat{R} = \bar{y}/\bar{x}$, the difference between (3.2) and (3.1) can be written as

$$\hat{\bar{Y}}^c_{\mathrm{RS}}(j) - \hat{\bar{Y}}^c_{\mathrm{RS}} =$$

$$\begin{cases} -\hat{R}\dfrac{(x_j - \bar{x}')}{(m-1)} - \dfrac{\bar{x}'(j)}{\bar{x}(j)}\dfrac{(y_j - \hat{R}x_j)}{(n-1)} & \text{if } j \in s_2 \\ -\hat{R}\dfrac{(x_j - \bar{x}')}{(m-1)} & \text{if } j \in (s_1 - s_2). \end{cases} \tag{3.3}$$

Expression (3.3) is exactly the same as reported by Rao and Sitter (1995). Assuming that $\bar{x}'(j)/\bar{x}(j) \approx \bar{x}'/\bar{x}$, then the approximate jackknife estimator of variance is given by

$$\hat{V}_{\mathrm{JACK}}(\hat{\bar{Y}}^c_{\mathrm{RS}}) \approx \left(\frac{\bar{x}'}{\bar{x}}\right)^2 \sum_{i\in s_2} \frac{(y_i - \hat{R}x_i)^2}{n(n-1)} + 2\left(\frac{\bar{x}'}{\bar{x}}\right)\hat{R}\sum_{j\in s_2}\frac{(x_j - \bar{x}')(y_j - \hat{R}x_j)}{n-1} + \hat{R}^2\sum_{j\in s_1}\frac{(x_j - \bar{x}')^2}{m(m-1)}.$$

Thus, the Rao and Sitter (1995) estimator is a special case of the proposed jackknife technique.

Case 3.2: Sitter (1997)

In Case 3.1, if we consider $q_{2i} = 1$, then the calibrated estimator under SRSWOR becomes

$$\hat{\bar{Y}}^c_{lr} = \bar{y} + b^*(\bar{x}' - \bar{x}), \tag{3.4}$$

where $b^* = \sum_{i\in s_2} x_i y_i / \sum_{i\in s_2} x_i^2$ denotes an estimator of the regression coefficient β that is slightly different from the one considered by Sitter (1997). The jackknife mechanism takes the form

$$\hat{\bar{Y}}^c_{lr}(j) = \begin{cases} \dfrac{n\bar{y} - y_j}{n-1} + \left\{ b^* + \dfrac{x_j(y_j - b^* x_j)}{\sum_{i\in s_2} x_i^2 - x_j^2} \right\} \left\{ \dfrac{m\bar{x}' - x_j}{m-1} - \dfrac{n\bar{x} - x_j}{n-1} \right\} & \text{if } j \in s_2 \\ \bar{y} + b^*\left\{ \dfrac{m\bar{x}' - x_j}{m-1} - \bar{x} \right\} & \text{if } j \in (s_1 - s_2). \end{cases} \tag{3.5}$$

If we set $d_j^* = (y_j - \bar{y}) - b^*(x_j - \bar{x})$, $a_j^* = x_j\{\bar{x}(j) - \bar{x}'(j)\}/K$, and $k_j^* = x_j^2/K$, where $K = (n-1)s_x^2 + n\bar{x}^2$, then the difference between (3.5) and (3.4) can be written as

$$\hat{\bar{Y}}^c_{lr}(j) - \hat{\bar{Y}}^c_{lr} =$$

$$\begin{cases} -b^*\dfrac{(x_j - \bar{x}')}{(m-1)} - \dfrac{\dfrac{d_j^*}{(n-1)}}{\left[1 + \dfrac{a_j^*}{(1-k_j^*)}\right]} & \text{if } j \in s_2 \\ -b^*\dfrac{(x_j - \bar{x}')}{(m-1)} & \text{if } j \in (s_1 - s_2) \end{cases}$$

which is similar to the expression reported by Sitter (1997).

Case 3.3: Raj (1965)

In order to consider this case, we assume that the initial sample s_1 of size m is selected with replacement according to probabilities p_i proportional to z_i, $i = 1, 2, \ldots, N$. Information on the auxiliary variable X is collected on this first-phase sample, s_1. The second-phase sample, specified to be of size n, is a subsample of s_1 selected without replacement using equal probabilities. It is for s_2 that information on Y is collected. Under this sampling scheme, $d_{1i} = 1/\pi_{1i} = 1/(mp_i)$ and $d_{2i} = m/n$. Thus, $w_{1i}^o = (1/p_i)/\sum_{i\in s_1}(1/p_i)$ and $w_{2i}^o = (1/p_i)/\sum_{i\in s_2}(1/p_i)$. Note also that for this scheme, $\hat{\bar{X}}_1^c = \hat{\bar{X}}_1^o$; thus no first-phase calibration is made. If $q_{2i} = 1/x_i$, then the calibrated estimator $\hat{\bar{Y}}^c$ becomes

$$\hat{\bar{Y}}^c_{\mathrm{Raj}} = \hat{\bar{Y}}_2^o(\hat{\bar{X}}_1^o / \hat{\bar{X}}_2^o), \tag{3.6}$$

where $\hat{\bar{Y}}_2^o = \sum_{i\in s_2}(y_i/p_i)/\sum_{i\in s_2}(1/p_i)$, $\hat{\bar{X}}_2^o = \sum_{i\in s_2}(x_i/p_i)/\sum_{i\in s_2}(1/p_i)$, and $\hat{\bar{X}}_1^o = \sum_{i\in s_1}(x_i/p_i)/\sum_{i\in s_1}(1/p_i)$. Thus, alternatively $\hat{\bar{Y}}^c_{\mathrm{Raj}} = \{\sum_{i\in s_2}(y_i/p_i)\sum_{i\in s_1}(x_i/p_i)\}/\{\sum_{i\in s_2}(x_i/p_i)\sum_{i\in s_1}(1/p_i)\}$.

Under the sampling scheme described above, the jackknife estimator of population mean is

$$\hat{\bar{Y}}^c_{\mathrm{Raj}}(j) = \begin{cases} \hat{\bar{Y}}_2^o(j)\dfrac{\hat{\bar{X}}_1^o(j)}{\hat{\bar{X}}_2^o(j)} & \text{if } j \in s_2 \\ \hat{\bar{Y}}_2^o\dfrac{\hat{\bar{X}}_1^o(j)}{\hat{\bar{X}}_2^o} & \text{if } j \in (s_1 - s_2) \end{cases} \tag{3.7}$$

where

$$\hat{\bar{Y}}_2^o(j) = \frac{\sum_{i\in s_2}(y_i/p_i)}{\sum_{i\in s_2}(1/p_i)} + \frac{(1/p_j)\Big/\sum_{i\in s_2}(1/p_i)}{1 - \dfrac{(1/p_j)}{\sum_{i\in s_2}(1/p_i)}}\left\{\frac{\sum_{i\in s_2}(y_i/p_i)}{\sum_{i\in s_2}(1/p_i)} - y_j\right\},$$

and $\hat{\bar{X}}_2^o(j)$ and $\hat{\bar{X}}_1^o(j)$ are defined analogously. If $\hat{R} = \hat{\bar{Y}}_2^o / \hat{\bar{X}}_2^o$ and $w_{2j}^o = (1/p_j)/\sum_{i\in s_2}(1/p_i)$, the difference between (3.7) and (3.6) can easily be written as

$$\hat{\bar{Y}}_{\text{Raj}}^c(j) - \hat{\bar{Y}}_{\text{Raj}}^c = \begin{cases} -w_{2j}^o \dfrac{\hat{\bar{X}}_1^o(j)}{\hat{\bar{X}}_2^o(j)}(y_j - \hat{R}x_j) \\ \qquad + \hat{R}\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_1^o\} & \text{if } j \in s_2 \\ \hat{R}\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_1^o\} & \text{if } j \in (s_1 - s_2). \end{cases}$$

Thus, the jackknife estimator of variance of the estimator $\hat{\bar{Y}}_{\text{Raj}}^c$ is given by

$$\hat{V}_{\text{JACK}}(\hat{\bar{Y}}_{\text{Raj}}^c) = \frac{m-1}{m}\Bigg[\sum_{j\in s_2}(w_{2j}^o)^2 \frac{\hat{\bar{X}}_1^o(j)^2}{\hat{\bar{X}}_2^o(j)}(y_j - \hat{R}x_j)^2 + \hat{R}^2\sum_{j\in s_1}\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_1^o\}^2 - 2\hat{R}\sum_{j\in s_2} w_{2j}^o \frac{\hat{\bar{X}}_1^o(j)}{\hat{\bar{X}}_2^o(j)}(y_j - \hat{R}x_j)\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_1^o\}\Bigg].$$

Following Rao and Sitter (1995), if we assume $\hat{\bar{X}}_1^o(j)/\hat{\bar{X}}_2^o(j) \approx \hat{\bar{X}}_1^o/\hat{\bar{X}}_2^o$, then the jackknife estimator of variance of $\hat{\bar{Y}}_{\text{Raj}}^c$ takes the form

$$\hat{V}_{\text{JACK}}(\hat{\bar{Y}}_{\text{Raj}}^c) \approx \frac{m-1}{m}\Bigg[\{\hat{\bar{X}}_1^o/\hat{\bar{X}}_2^o\}^2\sum_{j\in s_2}(w_{2j}^o)^2(y_j - \hat{R}x_j)^2 + \hat{R}^2\sum_{j\in s_1}\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_1^o\}^2 - 2\hat{R}\{\hat{\bar{X}}_1^o/\hat{\bar{X}}_2^o\}\sum_{j\in s_2} w_{2j}^o(y_j - \hat{R}x_j)\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_1^o\}\Bigg].$$

Case 3.4: Srivenkataramana and Tracy (1989)

In order to consider this case, as in Raj (1965), we assume that the initial sample s_1 of size m is selected with replacement according to probabilities proportional to z_i. However, the subsample, s_2, of n units is now selected with replacement using probabilities proportional to x_i/z_i. As a result, $w_{1i}^o = (1/z_i)/\sum_{i\in s_1}(1/z_i)$ and $w_{2i}^o = (1/x_i)/\sum_{i\in s_2}(1/x_i)$. Similar to Raj (1965), no first-phase calibration is made; thus $\hat{\bar{X}}_1^c = \hat{\bar{X}}_1^o$. Hence, if $q_{2i} = 1/x_i$, then the calibrated estimator $\hat{\bar{Y}}^c$ is

$$\hat{\bar{Y}}_{\text{ST}}^c = \hat{\bar{Y}}_2^o(\hat{\bar{X}}_1^o/\hat{\bar{X}}_2^o), \tag{3.8}$$

where $\hat{\bar{Y}}_2^o = \sum_{i\in s_2}(y_i/x_i)/\sum_{i\in s_2}(1/x_i)$, $\hat{\bar{X}}_2^o = n/\sum_{i\in s_2}(1/x_i)$, and $\hat{\bar{X}}_1^o = \sum_{i\in s_1}(x_i/z_i)/\sum_{i\in s_1}(1/z_i)$. Thus, alternatively $\hat{\bar{Y}}_{\text{ST}}^c = \{\sum_{i\in s_2}(y_i/x_i)\sum_{i\in s_1}(x_i/z_i)\}/\{n\sum_{i\in s_1}(1/z_i)\}$.

Under the sampling scheme described above, the jackknife estimator of population mean is

$$\hat{\bar{Y}}_{\text{ST}}^c(j) = \begin{cases} \hat{\bar{Y}}_2^o(j)\{\hat{\bar{X}}_1^o(j)/\hat{\bar{X}}_2^o(j)\} & \text{if } j \in s_2 \\ \hat{\bar{Y}}_2^o\{\hat{\bar{X}}_1^o(j)/\hat{\bar{X}}_2^o\} & \text{if } j \in (s_1 - s_2) \end{cases} \tag{3.9}$$

where

$$\hat{\bar{Y}}_2^o(j) = \frac{\sum_{i\in s_2}(y_i/x_i)}{\sum_{i\in s_2}(1/x_i)} + \frac{1}{x_j\sum_{i\in s_2}(1/x_i) - 1}\left\{\frac{\sum_{i\in s_2}(y_i/x_i)}{\sum_{i\in s_2}(1/x_i)} - y_j\right\}.$$

The terms $\hat{\bar{X}}_2^o(j)$ and $\hat{\bar{X}}_1^o(j)$ are defined similarly; that is

$$\hat{\bar{X}}_2^o(j) = \frac{n}{\sum_{i\in s_2}(1/x_i)} + \frac{1}{x_j\sum_{i\in s_2}(1/x_i) - 1}\left\{\frac{n}{\sum_{i\in s_2}(1/x_i)} - x_j\right\},$$

while $\hat{\bar{X}}_1^o(j)$ can be written as

$$\hat{\bar{X}}_1^o(j) = \frac{\sum_{i\in s_1}(x_i/z_i)}{\sum_{i\in s_1}(1/z_i)} + \frac{1}{x_j\sum_{i\in s_1}(1/x_i) - 1}\left\{\frac{\sum_{i\in s_1}(y_i/x_i)}{\sum_{i\in s_1}(1/x_i)} - x_j\right\}.$$

If $\hat{R} = \sum_{i\in s_2}(y_i/x_i)/n$ and $w_{2j}^o = (1/x_j)/\sum_{i\in s_2}(1/x_i)$, the difference between (3.9) and (3.8) is given by

$$\hat{\bar{Y}}_{\text{ST}}^c(j) - \hat{\bar{Y}}_{\text{ST}}^c = \begin{cases} -w_{2j}^o \dfrac{\hat{\bar{X}}_1^o(j)}{\hat{\bar{X}}_2^o(j)}(y_j - \hat{R}x_j) + \hat{R}\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_1^o\} & \text{if } j \in s_2 \\ \hat{R}\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_1^o\} & \text{if } j \in (s_1 - s_2). \end{cases}$$

Following Rao and Sitter (1995), if we assume $\hat{\bar{X}}_1^o(j)/\hat{\bar{X}}_2^o(j) \approx \hat{\bar{X}}_1^o/\hat{\bar{X}}_2^o$, then the jackknife estimator of variance of $\hat{\bar{Y}}_{\text{ST}}^c$ takes the form

$$\hat{V}_{\text{JACK}}(\hat{\bar{Y}}_{\text{ST}}^c) \approx \frac{m-1}{m}\Bigg[\{\hat{\bar{X}}_1^o/\hat{\bar{X}}_2^o\}^2\sum_{j\in s_2}(w_{2j}^o)^2(y_j - \hat{R}x_j)^2 + \hat{R}^2\sum_{j\in s_1}\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_1^o\}^2 - 2\hat{R}\{\hat{\bar{X}}_1^o/\hat{\bar{X}}_2^o\}\sum_{j\in s_2} w_{2j}^o(y_j - \hat{R}x_j)\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_1^o\}\Bigg].$$

Case 3.5: Chand (1975)

In order to consider this case, the first-phase sample s_1 of size m is selected using SRSWOR, and both auxiliary

variables Z and X are observed on the chosen units. The subsample, s_2, of n units is also selected using SRSWOR. Obviously, $d_{1i} = N/m$ and $d_{2i} = m/n$, so that $w_{1i}^o = 1/m$ and $w_{2i}^o = 1/n$. If $q_{1i} = 1/z_i$ and $q_{2i} = 1/x_i$, then the calibrated estimator $\hat{\bar{Y}}^c$ becomes

$$\hat{\bar{Y}}_{\text{Ch}}^c = \bar{y}(\bar{x}'/\bar{x})(\bar{Z}/\bar{z}'), \tag{3.10}$$

where

$$\bar{y} = \sum_{i\in s_2} y_i/n,\ \bar{x} = \sum_{i\in s_2} x_i/n,\ \bar{x}' = \sum_{i\in s_1} x_i/m,$$

and $\bar{z}' = \sum_{i\in s_1} z_i/m$. The jackknife estimator of $\bar{Y}$ is

$$\hat{\bar{Y}}_{\text{Ch}}^c(j) = \begin{cases} \bar{y}(j)\dfrac{\bar{x}'(j)}{\bar{x}(j)}\dfrac{\bar{Z}}{\bar{z}'(j)} & \text{if } j\in s_2 \\ \bar{y}(j)\dfrac{\bar{x}'(j)}{\bar{x}}\dfrac{\bar{Z}}{\bar{z}'(j)} & \text{if } j\in(s_1-s_2) \end{cases} \tag{3.11}$$

where $\bar{y}(j) = (n\bar{y}-y_j)/(n-1)$, $\bar{x}(j) = (n\bar{x}-x_j)/(n-1)$, $\bar{x}'(j) = (m\bar{x}'-x_j)/(m-1)$, and finally $\bar{z}'(j) = (m\bar{z}'-z_j)/(m-1)$. If we let $\hat{R}_1 = \bar{x}'/\bar{z}'$ (an estimator of $R_1 = \bar{X}/\bar{Z}$) and $\hat{R}_2 = \bar{y}/\bar{x}$ (an estimator of $R_2 = \bar{Y}/\bar{X}$), and similarly, let $\hat{R}_1(j) = \bar{x}'(j)/\bar{z}'(j)$ and $\hat{R}_2(j) = \bar{y}(j)/\bar{x}(j)$, the difference between (3.11) and (3.10) can be written as

$$\hat{\bar{Y}}_{\text{Ch}}^c(j) - \hat{\bar{Y}}_{\text{Ch}}^c = \begin{cases} \varepsilon_2(j)+\hat{R}_2\varepsilon_1(j)+\hat{R}_2(j)d_2(j)+\hat{R}_2\delta_2(j) & \text{if } j\in s_2 \\ \hat{R}_2\varepsilon_1(j) & \text{if } j\in(s_1-s_2) \end{cases} \tag{3.12}$$

where we can write in (3.12) that $\varepsilon_2(j) = \{\bar{y}(j)-\bar{y}\} - \hat{R}_2(j)\{\bar{x}(j)-\bar{x}\} - \hat{R}_1(j)\hat{R}_2(j)\{\bar{z}'(j)-\bar{Z}\}$, $d_2(j) = \{\bar{x}'(j)-\bar{x}'\}$, $\delta_2(j) = \{\bar{x}(j)-\bar{x}'(j)\} - \hat{R}_1(j)\{\bar{Z}-\bar{z}'(j)\} - \hat{R}_1\{\bar{Z}-\bar{z}'\}$, and finally that the term $\varepsilon_1(j) = \{\bar{x}'(j)-\bar{x}'\} - \hat{R}_1(j)\{\bar{z}'(j)-\bar{Z}\}$. Thus the jackknife estimator of variance of the estimator $\hat{\bar{Y}}_{\text{Ch}}^c$ is given by

$$\begin{aligned}\hat{V}_{\text{JACK}}(\hat{\bar{Y}}_{\text{Ch}}^c) = {}& \{(m-1)/m\}\Big[\sum_{j\in s_2}\varepsilon_2^2(j) + \sum_{j\in s_2}\hat{R}_2^2(j)d_2^2(j) \\ & + \hat{R}_2^2\sum_{j\in s_2}\delta_2(j)\{\delta_2(j)+2\varepsilon_1(j)\} \\ & + 2\hat{R}_2\sum_{j\in s_2}\varepsilon_1(j)\varepsilon_2(j) \\ & + 2\hat{R}_2\sum_{j\in s_2}\hat{R}_2(j)d_2(j)\{\varepsilon_1(j)+\delta_2(j)\} \\ & + \hat{R}_2^2\sum_{j\in s_1}\varepsilon_1^2(j)\Big].\end{aligned}$$

Case 3.6: Ahmed (1997)

Consider the same sample design as in Case 3.5. Rather than $q_{1i}=1/z_i$ and $q_{2i}=1/x_i$ as in Chand (1975), we set $q_{1i}=q_{2i}=1$, and $q_{2i}=1/x_i$, then the calibrated estimator reduces to

$$\hat{\bar{Y}}_{\text{Chlr}}^c = \bar{y} + b_2^*(\bar{x}'-\bar{x}) + b_1^*b_2^*(\bar{Z}-\bar{z}'), \tag{3.13}$$

where $b_2^* = \sum_{i\in s_2}x_iy_i/\sum_{i\in s_2}x_i^2$ and $b_1^* = \sum_{i\in s_1}x_iz_i/\sum_{i\in s_1}z_i^2$. Note that (3.13) is a chain regression type estimator similar to Ahmed (1997). Letting $b_2^*(j) = b_2^* + \{x_j(y_j - b_2^*x_j)/(x_j^2-\sum_{i\in s_2}x_i^2)\}$ and $b_1^*(j) = b_1^* + \{z_j(x_j-b_1^*z_j)/(z_j^2-\sum_{i\in s_1}z_i^2)\}$, after jackknifing the estimator $\hat{\bar{Y}}_{\text{Chlr}}^c$ becomes

$$\hat{\bar{Y}}_{\text{Chlr}}^c(j) = \begin{cases} \bar{y}(j) + b_2^*(j)\{\bar{x}'(j)-\bar{x}(j)\} \\ \quad + b_1^*(j)b_2^*(j)\{\bar{Z}-\bar{z}'(j)\} & \text{if } j\in s_2 \\ \bar{y} + b_2^*\{\bar{x}'(j)-\bar{x}\} \\ \quad + b_1^*(j)b_2^*\{\bar{Z}-\bar{z}'(j)\} & \text{if } j\in(s_1-s_2). \end{cases} \tag{3.14}$$

The difference between (3.14) and (3.13) can be written as

$$\hat{\bar{Y}}_{\text{Chlr}}^c(j) - \hat{\bar{Y}}_{\text{Chlr}}^c = \begin{cases} \varepsilon_2(j)+b_2^*\varepsilon_1(j)+b_2^*(j)d_2(j)+b_2^*\delta_2(j) & \text{if } j\in s_2 \\ b_2^*\varepsilon_1(j) & \text{if } j\in(s_1-s_2) \end{cases} \tag{3.15}$$

where we can write in (3.15) that $\varepsilon_2(j) = \{\bar{y}(j) - \bar{y}\} - b_2^*(j)\{\bar{x}(j)-\bar{x}\} - b_1^*(j)\,b_2^*(j)\{\bar{z}'(j)-\bar{Z}\}$, $d_2(j) = \{\bar{x}'(j)-\bar{x}'\}$, $\delta_2(j) = \{\bar{x}(j)-\bar{x}'(j)\} - b_1^*(j)\{\bar{Z}-\bar{z}'(j)\} - b_1^*\{\bar{Z}-\bar{z}'\}$, and finally that the term $\varepsilon_1(j) = \{\bar{x}'(j) - \bar{x}'\} - b_1^*(j)\{\bar{z}'(j)-\bar{Z}\}$. Thus the jackknife estimator of variance of the estimator $\hat{\bar{Y}}_{\text{Chlr}}^c$ is given by

$$\begin{aligned}\hat{V}_{\text{JACK}}(\hat{\bar{Y}}_{\text{Chlr}}^c) = {}& \{(m-1)/m\}\Big[\sum_{j\in s_2}\varepsilon_2^2(j) + \sum_{j\in s_2}\{b_2^*(j)\}^2d_2^2(j) \\ & + \{b_2^*(j)\}^2\sum_{j\in s_2}\delta_2(j)\{\delta_2(j)+2\varepsilon_1(j)\} \\ & + 2b_2^*\sum_{j\in s_2}\varepsilon_1(j)\varepsilon_2(j) \\ & + 2b_2^*\sum_{j\in s_2}b_2^*(j)d_2(j)\{\varepsilon_1(j)+\delta_2(j)\} \\ & + \{b_2^*\}^2\sum_{j\in s_1}\varepsilon_1^2(j)\Big].\end{aligned}$$

4. Simulation study

In this section, we present the results of simulation studies designed to investigate the performance of the proposed jackknife procedure for estimating the variance of four of the two-phase estimators of population mean

presented in Section 3. Specifically, we consider the Rao and Sitter (1995) ratio-type estimator, the Sitter (1997) regression-type estimator, the Chand (1975) chain ratio-type estimator, and the Ahmed (1997) chain regression-type estimator. Initially, we describe and report the results of simulations that were conducted for the Sitter and Rao (1995) and Sitter (1997) estimators. This is followed by a discussion and summary of similar simulations on the Chand (1975) and Ahmed (1997) estimators. Unlike the case for the ratio and regression estimators, since complete information on a second auxiliary variable Z is required for the entire population in order to apply the two chain estimators, the simulations that were conducted for these two estimators are somewhat more complicated than those performed for the ratio and regression estimators.

4.1 Simulation study: Rao and Sitter (1995) and Sitter (1997)

For purposes of the first set of simulations, we assume that a first-phase sample of m units is selected from a population of N units, and only the auxiliary variable X is measured. From the first-phase sample of m units, we then select a second-phase sample of n units by SRSWOR in which both the study variable, Y, and the auxiliary variable, X, are measured.

We began by creating a population of N units consisting of (X_i, Y_i) pairs using the model

$$Y_i = \beta X_i + \sqrt{X_i^g}\,\varepsilon_i,$$

with $\beta = 10$. Initially, we set $g = 0$ and $N = 500$. For each i, $i = 1, ..., N$, we generated X_i from a gamma distribution with a shape parameter of 3.1 and a scale parameter of one, and ε_i from a standard normal. From the resulting population of (X_i, Y_i) pairs, we selected 1,000 first-phase sample of $m = 100$ units, and from each of these samples, we selected 10,000 second-phase samples of $n = 20$ units.

Under the sampling scheme used here, Rao and Sitter (1995) proposed the ratio estimator

$$\hat{\bar{Y}}_{\mathrm{RS}}^{c} = \bar{y}(\bar{x}'/\bar{x}), \tag{4.1}$$

which has approximate variance

$$V(\hat{\bar{Y}}_{\mathrm{RS}}^{c}) = (n^{-1} - m^{-1})S_d^2 + (m^{-1} - N^{-1})S_y^2,$$

where

$$S_d^2 = (N-1)^{-1}\sum_{i=1}^{N}[(Y_i - \bar{Y}) - R(X_i - \bar{X})]^2$$

and

$$S_y^2 = (N-1)^{-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2,$$

with $\bar{Y} = \sum_{i=1}^{N} Y_i/N$, $\bar{X} = \sum_{i=1}^{N} X_i/N$, and $R = \bar{Y}/\bar{X}$. For the t^{th} second phase sample ($t = 1, \ldots, 10{,}000$) drawn from the k^{th} first phase sample ($k = 1, \ldots, 1{,}000$), we computed the usual estimator of variance

$$\hat{V}[(\hat{\bar{Y}}_{\mathrm{RS}}^{c}(t|k)] = \left(\frac{1}{n} - \frac{1}{m}\right)s_{d(t|k)}^2 + \left(\frac{1}{m} - \frac{1}{N}\right)s_{y(t|k)}^2, \tag{4.2}$$

where the sample variances are

$$s_{d(t|k)}^2 = (n-1)^{-1}\sum_{i=1}^{n}[(y_{i(t|k)} - \bar{y}_{(t|k)}) - r_{(t|k)}(x_{i(t|k)} - \bar{x}_{(t|k)})]^2$$

and

$$s_{y(t|k)}^2 = (n-1)^{-1}\sum_{i=1}^{n}(y_{i(t|k)} - \bar{y}_{(t|k)})^2$$

with $\bar{y}_{(t|k)} = \sum_{i=1}^{n} y_{i(t|k)}/n$ and $\bar{x}_{(t|k)} = \sum_{i=1}^{n} x_{i(t|k)}/n$. In addition, $r_{(t|k)} = \bar{y}_{(t|k)}/\bar{x}_{(t|k)}$. We also computed the jackknife estimator of variance

$$\hat{V}_{\mathrm{JACK}}[(\hat{\bar{Y}}_{\mathrm{RS}}^{c}(t|k)] =$$

$$\frac{m-1}{m}\sum_{j=1}^{m}\left[\bar{y}_{(t|k)}(j)\frac{\bar{x}'_{(t|k)}(j)}{\bar{x}_{(t|k)}(j)} - \bar{y}_{(t|k)}\frac{\bar{x}'_{(t|k)}}{\bar{x}_{(t|k)}}\right]^2, \tag{4.3}$$

and the ratio of estimated variances

$$\mathrm{RV}(t|k) = \hat{V}[(\hat{\bar{Y}}_{\mathrm{RS}}^{c}(t|k)]/\hat{V}_{\mathrm{JACK}}[(\hat{\bar{Y}}_{\mathrm{RS}}^{c}(t|k)].$$

We then computed the average of the $\mathrm{RV}(t|k)$ over all k and t, which is given by

$$\mathrm{RV} = \frac{1}{10{,}000{,}000}\sum_{k=1}^{1{,}000}\sum_{t=1}^{10{,}000}\mathrm{RV}(t|k).$$

We also determined empirical estimates of the biases in (4.2) and (4.3) by computing

$$\mathrm{EBU} = \frac{1}{10{,}000{,}000}\sum_{k=1}^{1{,}000}\sum_{t=1}^{10{,}000}\{\hat{V}[(\hat{\bar{Y}}_{\mathrm{RS}}^{c}(t|k)] - V(\hat{\bar{Y}}_{\mathrm{RS}}^{c})\},$$

and

$$\mathrm{EBJ} = \frac{1}{10{,}000{,}000}\sum_{k=1}^{1{,}000}\sum_{t=1}^{10{,}000}\{\hat{V}_{\mathrm{JACK}}[(\hat{\bar{Y}}_{\mathrm{RS}}^{c}(t|k)] - V(\hat{\bar{Y}}_{\mathrm{RS}}^{c})\}.$$

Note that the estimator given in (4.2) is unbiased. Finally, we calculated the relative efficiency of the usual estimator of variance to the jackknife estimator according to

$$\mathrm{RE} = \left(\frac{\sum_{k=1}^{1{,}000}\sum_{t=1}^{10{,}000}\{\hat{V}[(\hat{\bar{Y}}_{\mathrm{RS}}^{c}(t|k)] - V(\hat{\bar{Y}}_{\mathrm{RS}}^{c})\}^2}{\sum_{k=1}^{1{,}000}\sum_{t=1}^{10{,}000}\{\hat{V}_{\mathrm{JACK}}[(\hat{\bar{Y}}_{\mathrm{RS}}^{c}(t|k)] - V(\hat{\bar{Y}}_{\mathrm{RS}}^{c})\}^2}\right).$$

Using the same generated population of $N = 500$, we repeated the simulation; however we used $m = 400$ and $n = 80$ instead. We then created four additional populations of size $N = 500$ using $g = 0.5$, 1.0, 1.5, and 2.0. For each of these four populations, we repeated the two simulations described above where in the first simulation, $m = 100$ with $n = 20$, and in the second simulation, $m = 400$ and $n = 80$. Finally, to study the effect of population size, we then repeated all the simulations based on the different values of g, m, and n when $N = 500$ for three additional values of N, namely 5,000, 50,000, and 500,000. The results obtained for RV, EBU, EBJ, and RE for each of these simulations are presented in Table 1.

The results for RE in Table 1 suggest that as the population size N tends to infinity (as considered by Rao and Sitter 1995), the jackknife estimator of variance remains more efficient than the usual unbiased estimator of variance. It is also the case for very large N that the values for RV tend to one. However, considering the cases where $N = 500$, if the population size is relatively small, not only are the values for RV noticeably smaller than one, but the jackknife estimator of variance seems to be significantly biased. In addition, the jackknife estimator appears to be much less efficient than the usual unbiased estimator of variance, especially when m and n are large. Of note here is the fact that Rao and Sitter (1995) and Sitter (1997) state that it is not clear how to fix the finite population correction factors in the jackknife estimator of variance in two-phase sampling. This would seem to be an area where further research could be fruitful, since it would appear that when the population size is small, it might be worthwhile to adjust the finite population correction factors instead of directly applying the jackknife technique according to the approach proposed here. Note that Kim *et al.* (2006) have incorporated a finite population correction factor in a special case.

Table 1
Comparison of the jackknife and usual estimators of variance of the ratio estimator of the population mean when β = 10 and the auxiliary variable, *X*, follows a gamma distribution with a shape parameter of 3.1 and a scale parameter of one

N	*m*	*n*	*g*	RV	EBU	EBJ	RE
500	100	20	0.0	0.801	0.006	0.542	1.521
			0.5	0.800	0.010	0.579	1.310
			1.0	0.805	-0.071	0.561	1.267
			1.5	0.816	-0.358	0.575	1.149
			2.0	0.840	-0.720	1.777	0.935
5,000	100	20	0.0	0.979	-0.028	0.042	4.015
			0.5	0.976	0.007	0.096	3.709
			1.0	0.965	0.023	0.172	3.210
			1.5	0.936	-0.073	0.337	1.308
			2.0	0.916	-1.103	0.493	0.967
50,000	100	20	0.0	1.001	-0.002	0.003	6.241
			0.5	0.998	0.107	0.126	4.936
			1.0	0.981	0.101	0.196	2.965
			1.5	0.937	-0.211	0.167	1.558
			2.0	0.924	-0.355	0.940	1.005
500,000	100	20	0.0	1.001	-0.057	-0.054	4.730
			0.5	0.999	0.014	0.024	4.669
			1.0	0.993	0.185	0.229	3.223
			1.5	0.940	-0.235	0.122	1.420
			2.0	0.907	-1.054	0.530	1.009
500	400	80	0.0	0.214	0.000	0.520	0.002
			0.5	0.237	-0.001	0.523	0.002
			1.0	0.320	0.000	0.544	0.006
			1.5	0.530	-0.001	0.616	0.066
			2.0	0.733	-0.012	1.091	0.452
5,000	400	80	0.0	0.919	-0.003	0.061	2.687
			0.5	0.920	-0.001	0.064	2.505
			1.0	0.922	0.003	0.077	2.058
			1.5	0.930	-0.028	0.077	1.372
			2.0	0.940	-0.089	0.184	1.088
50,000	400	80	0.0	0.991	-0.008	-0.001	4.550
			0.5	0.991	0.004	0.012	5.276
			1.0	0.991	0.000	0.009	4.163
			1.5	0.980	-0.024	-0.001	1.777
			2.0	0.967	-0.171	-0.040	1.099
500,000	400	80	0.0	1.000	0.009	0.009	5.501
			0.5	0.999	0.001	0.001	5.180
			1.0	0.993	-0.001	0.006	3.852
			1.5	0.992	-0.022	-0.018	1.809
			2.0	0.971	-0.179	-0.079	1.136

We also considered the Sitter (1997) regression estimator, and repeated the entire simulation study that was performed using the ratio estimator in (4.1). Specifically, rather than (4.1), we made use of the estimator

$$\hat{\bar{Y}}_S^c = \bar{y} + b^*(\bar{x}' - \bar{x}), \quad (4.4)$$

which has approximate variance

$$V(\hat{\bar{Y}}_S^c) = (n^{-1} - m^{-1})S_d^2 + (m^{-1} - N^{-1})S_y^2, \quad (4.5)$$

where

$$S_d^2 = (N-1)^{-1}\sum_{i=1}^{N}[(Y_i - \bar{Y}) - \beta_{\text{POP}}(X_i - \bar{X})]^2$$

with

$$\beta_{\text{POP}} = \sum_{i=1}^{N} X_i Y_i / \sum_{i=1}^{N} X_i^2.$$

For each different combination of N, g, m, and n used in the simulation study, we computed

$$\hat{V}[(\hat{\bar{Y}}_S^c(t|k)] = (n^{-1} - m^{-1})s_{d(t|k)}^2 + (m^{-1} - N^{-1})s_{y(t|k)}^2, \quad (4.6)$$

for the t^{th} second phase sample drawn from the k^{th} first phase sample, where the sample variance

$$s_{d(t|k)}^2 = (n-1)^{-1}\sum_{i=1}^{n}[(y_{i(t|k)} - \bar{y}_{(t|k)}) - b_{(t|k)}^*(x_{i(t|k)} - \bar{x}_{(t|k)})]^2.$$

We also computed the jackknife estimator of variance

$$\hat{V}_{\text{JACK}}[(\hat{\bar{Y}}_S^c(t|k)] =$$

$$\frac{m-1}{m}\sum_{j=1}^{m}[\bar{y}_{(t|k)}(j) + b_{(t|k)}^*(j)\{\bar{x}'_{(t|k)}(j) - \bar{x}_{(t|k)}(j)\} - \{\bar{y} + b^*(\bar{x}' - \bar{x})\}]^2. \quad (4.7)$$

For each different combination of N, g, m, and n, equations (4.5) through (4.7) were used to compute values for RV, EBU, EBJ, and RE analogous to those given in Table 1 for the estimator in (4.1). The results obtained were extremely similar to those for the ratio estimator.

4.2 Simulation study: Chand (1975) and Ahmed (1997)

For purposes of the second set of simulations, we now assume that when the first-phase sample of m units is selected from the population of size N, information on two auxiliary variables X and Z is collected. When the second-phase sample of size n is selected from the first-phase sample, the study variable Y is measured, along with the two auxiliary variables X and Z. Note also that the auxiliary variable Z is assumed to be known for the entire population.

We began by creating a population of $N = 500$ units of (X_i, Z_i, Y_i) observations using

$$Y_i = \beta_1 X_i + \beta_2 Z_i + \varepsilon_i,$$

with $\beta_1 = 3.5$ and $\beta_2 = 2.5$. For each i, $i = 1, ..., N$, we generated X_i from a gamma distribution with a shape parameter of 2.2 and a scale parameter of one, Z_i from a gamma distribution with a shape parameter of 0.1 and a scale parameter of one, and ε_i from a standard normal. From the resulting population of (X_i, Z_i, Y_i) observations, we selected 1,000 first-phase sample of $m = 100$ units, and from each of these samples, we selected 10,000 second-phase samples of $n = 20$ units.

Following Chand (1975), a chain ratio estimator under two-phase sampling is given by

$$\hat{\bar{Y}}_{\text{Ch}}^c = \bar{y}(\bar{x}'/\bar{x})(\bar{Z}/\bar{z}'),$$

which has approximate variance

$$V(\hat{\bar{Y}}_{\text{Ch}}^c) = (n^{-1} - m^{-1})S_{d_2}^2 + (m^{-1} - N^{-1})S_{d_1}^2, \quad (4.8)$$

where

$$S_{d_2}^2 = (N-1)^{-1}\sum_{i=1}^{N}[(Y_i - \bar{Y}) - R_2(X_i - \bar{X})]^2$$

and

$$S_{d_1}^2 = (N-1)^{-1}\sum_{i=1}^{N}[(Y_i - \bar{Y}) - R_1(Z_i - \bar{Z})]^2$$

with

$$\bar{Y} = \sum_{i=1}^{N} Y_i/N,\ \bar{X} = \sum_{i=1}^{N} X_i/N,\ \bar{Z} = \sum_{i=1}^{N} Z_i/N,$$

$R_1 = \bar{Y}/\bar{Z}$, and $R_2 = \bar{Y}/\bar{X}$. In the simulation study, we computed

$$\hat{V}[(\hat{\bar{Y}}_{\text{Ch}}^c(t|k)] = (n^{-1} - m^{-1})s_{d_2(t|k)}^2 + (m^{-1} - N^{-1})s_{d_1(t|k)}^2, \quad (4.9)$$

for the t^{th} second phase sample drawn from the k^{th} first phase sample, where the sample variances

$$s_{d_2(t|k)}^2 = (n-1)^{-1}\sum_{i=1}^{n}[(y_{i(t|k)} - \bar{y}_{(t|k)}) - r_{2(t|k)}(x_{i(t|k)} - \bar{x}_{(t|k)})]^2$$

with

$$r_{2(t|k)} = \bar{y}_{(t|k)}/\bar{x}_{(t|k)}$$

and

$$s^2_{d_1(t|k)} = (n-1)^{-1}\sum_{i=1}^{n}[(y_{i(t|k)} - \bar{y}_{(t|k)}) - r_{1(t|k)}(z_{i(t|k)} - \bar{z}_{(t|k)})]^2$$

with $r_{1(t|k)} = \bar{y}_{(t|k)} / \bar{z}_{(t|k)}$. We also computed the jackknife estimator of variance

$$\hat{V}_{\text{JACK}}[(\hat{\bar{Y}}^c_{\text{Ch}}(t|k)] = \frac{m-1}{m}\sum_{j=1}^{m}\left[\bar{y}_{(t|k)}(j)\frac{\bar{x}'_{(t|k)}(j)}{\bar{x}_{(t|k)}(j)}\frac{\bar{Z}}{\bar{z}'_{(t|k)}(j)} - \bar{y}_{(t|k)}\frac{\bar{x}'_{(t|k)}}{\bar{x}_{(t|k)}}\frac{\bar{Z}}{\bar{z}'_{(t|k)}}\right]^2. \quad (4.10)$$

Using the same generated population of $N = 500$, we repeated the simulation; however we used $m = 400$ and $n = 80$ instead. We then created three additional populations of size $N = 500$ using $\beta_1 = 0.5$ with $\beta_2 = 0.5$, $\beta_1 = 3.5$ with $\beta_2 = 0.5$, and $\beta_1 = 0.5$ with $\beta_2 = 2.5$. For each of these three populations, we repeated the two simulations described above where in the first simulation, $m = 100$ with $n = 20$, and in the second simulation, $m = 400$ and $n = 80$. Finally, to study the effect of population size, we then repeated all the simulations based on the different values of β_1, β_2, m, and n when $N = 500$ for three additional values of N, namely 5,000, 50,000, and 500,000. For each different combination of N, β_1, β_2, m, and n, equations (4.8) through (4.10) were used to compute values for RV, EBU, EBJ, and RE analogous to those given in Table 1 for the estimator in (4.1). The results are provided in Table 2.

Generally speaking, the findings based on the results in Table 2 are similar to those arrived at for the estimators based on (4.1) and (4.4). In particular, the jackknife estimator of variance is more efficient than the usual estimator when the population size is sufficiently large. However, also of note is the fact that this efficiency seems to be related to the magnitude of the regression coefficients β_1 and β_2; that is, the jackknife estimator appears to achieve relatively greater efficiency for cases where the coefficient associated with the auxiliary variable X, is large relative to the analogous coefficient linked to Z.

Table 2
Comparison of the jackknife and usual estimators of variance of the chain ratio estimator of the population mean where the auxiliary variable, *X*, follows a gamma distribution with a shape parameter of 2.2 and a scale parameter of one, and the auxiliary variable, *Z*, follows a gamma distribution with a shape parameter of 0.1 and a scale parameter of one

m	*n*	β_1	β_2	*N*	RV	EBU	EBJ	RE
100	20	3.5	2.5	500	0.769	0.000	0.027	1.063
				5,000	0.831	-0.012	0.020	2.282
				50,000	0.818	-0.006	0.028	1.785
				500,000	0.852	0.001	0.036	1.993
100	20	0.5	0.5	500	0.911	-0.001	0.004	0.791
				5,000	0.943	-0.001	0.002	0.888
				50,000	0.948	0.000	0.003	0.896
				500,000	0.946	0.000	0.003	0.899
100	20	3.5	0.5	500	0.845	-0.001	0.015	1.674
				5,000	0.932	-0.011	0.000	3.632
				50,000	0.947	-0.005	0.004	3.221
				500,000	0.947	0.000	0.010	3.637
100	20	0.5	2.5	500	0.866	-0.001	0.009	0.668
				5,000	0.858	-0.003	0.008	0.775
				50,000	0.855	-0.001	0.010	0.670
				500,000	0.855	0.000	0.012	0.697
400	80	3.5	2.5	500	0.540	0.000	0.013	0.044
				5,000	0.780	-0.001	0.009	1.346
				50,000	0.819	0.000	0.008	1.878
				500,000	0.810	-0.001	0.006	1.953
400	80	0.5	0.5	500	0.817	0.000	0.003	0.254
				5,000	0.956	0.000	0.000	0.885
				50,000	0.973	0.000	0.001	0.946
				500,000	0.973	0.000	0.000	0.963
400	80	3.5	0.5	500	0.579	0.000	0.010	0.041
				5,000	0.907	-0.001	0.003	3.158
				50,000	0.954	0.000	0.002	3.845
				500,000	0.950	-0.001	0.001	4.853
400	80	0.5	2.5	500	0.787	0.000	0.004	0.222
				5,000	0.862	0.000	0.002	0.570
				50,000	0.873	0.000	0.003	0.698
				500,000	0.875	0.000	0.002	0.595

Finally, an analogous simulation study was performed using the regression estimator of Ahmed (1997). However, the populations were created using $\beta_1 = 10$ with $\beta_2 = 0.5$, $\beta_1 = 100$ with $\beta_2 = 0.5$, $\beta_1 = 0.5$ with $\beta_2 = 10$, and $\beta_1 = 10$ with $\beta_2 = 10$. As before when the estimators of Rao and Sitter (1995), Sitter (1997), and Chand (1975) were considered, provided that the population is sufficiently large, the jackknife estimator of variance seems to be more efficient than the usual estimator.

5. Conclusion and discussion

In this paper, the problem of estimating the variance of various estimators of the population mean in two-phase sampling has been considered by jackknifing the famous two-phase calibrated weights of Hidiroglou and Särndal (1995, 1998). Simulation studies based on ratio, regression, and chain-type estimators suggest that provided that the population size is large enough and the first and second-phase samples are relatively small, the jackknife estimator of variance is more efficient than the usual estimator of variance, regardless of the estimator for the population mean that is considered. For small populations, it might be worthwhile to adjust the finite population correction factors instead of directly applying the jackknife technique. This is an area where further research could be conducted.

Acknowledgements

This work was conducted while Sarjinder Singh was a postdoctoral fellow at Carleton University. The authors are grateful to the Associate Editor and the referees, whose comments greatly improved this manuscript. This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

Appendix

Derivation of the jackknife estimator in (2.1)

In this part of the appendix, we prove (2.1) for the jackknifed estimator of the population mean in two phase-sampling. First, note that $\hat{\beta}_1(j) = \hat{\beta}_1 + t_{1j}\, e_{1j}$ and $\hat{\beta}_2(j) = \hat{\beta}_2 + t_{2j} e_{2j}$, where $t_{1j} = q_{1j} w_{1j}^o z_j / (q_{1j} w_{1j}^o z_j^2 - \sum_{i \in s_1} q_{1i} w_{1i}^o z_i^2)$, $e_{1j} = x_j - \hat{\beta}_1 z_j$, $t_{2j} = q_{2j} w_{2j}^o x_j / (q_{2j} w_{2j}^o x_j^2 - \sum_{i \in s_1} q_{2i} w_{2i}^o x_i^2)$, and $e_{2j} = y_j - \hat{\beta}_2 x_j$. We also have $\hat{\bar{Z}}_1^o(j) = \hat{\bar{Z}}_1^o + h_{1j}(\hat{\bar{Z}}_1^o - z_j)$, $\hat{\bar{X}}_1^o(j) = \hat{\bar{X}}_1^o + h_{1j}(\hat{\bar{X}}_1^o - x_j)$, $\hat{\bar{X}}_2^o(j) = \hat{\bar{X}}_2^o + h_{2j}(\hat{\bar{X}}_2^o - x_j)$, and $\hat{\bar{Y}}_2^o(j) = \hat{\bar{Y}}_2^o + h_{2j}(\hat{\bar{Y}}_2^o - y_j)$, where $h_{1j} = w_{1j}^o / (1 - w_{1j}^o)$ and $h_{2j} = w_{2j}^o / (1 - w_{2j}^o)$.

Using these results, for $j \in s_2$, we have

$$\begin{aligned}\hat{\bar{Y}}^c(j) = {} & \hat{\bar{Y}}_2^o + \hat{\beta}_2(\hat{\bar{X}}_1^o - \hat{\bar{X}}_2^o) + \hat{\beta}_1\hat{\beta}_2(\bar{Z} - \hat{\bar{Z}}_1^o) \\ & + h_{2j}(\hat{\bar{Y}} - y_j) + t_{2j}\, e_{2j}(\hat{\bar{X}}_1^o - \hat{\bar{X}}_2^o) \\ & + \hat{\beta}_2\{h_{1j}(\hat{\bar{X}}_1^o - x_j) - h_{2j}(\hat{\bar{X}}_2^o - x_j)\} \\ & + t_{1j}\, e_{1j}\hat{\beta}_2(\bar{Z} - \hat{\bar{Z}}_1^o) - t_{1j}\, e_{1j}\hat{\beta}_2 h_{1j}(\hat{\bar{Z}}_1^o - z_j) \\ & + \hat{\beta}_1 t_{2j}\, e_{2j}(\bar{Z} - \hat{\bar{Z}}_1^o) - t_{2j}\, e_{2j}\hat{\beta}_1 h_{1j}(\hat{\bar{Z}}_1^o - z_j) \\ & - \hat{\beta}_1\hat{\beta}_2 h_{1j}(\hat{\bar{Z}}_1^o - z_j).\end{aligned}$$

Similarly, for $j \in (s_1 - s_2)$, we have

$$\begin{aligned}\hat{\bar{Y}}^c(j) = {} & \hat{\bar{Y}}_2^o + \hat{\beta}_2(\hat{\bar{X}}_1^o - \hat{\bar{X}}_2^o) + \hat{\beta}_1\hat{\beta}_2(\bar{Z} - \hat{\bar{Z}}_1^o) \\ & + \hat{\beta}_2 h_{1j}(\hat{\bar{X}}_1^o - x_j) + t_{1j}\, e_{1j}\hat{\beta}_2(\bar{Z} - \hat{\bar{Z}}_1^o) \\ & - t_{1j}\, e_{1j}\hat{\beta}_2 h_{1j}(\hat{\bar{Z}}_1^o - z_j) \\ & + \hat{\beta}_1\hat{\beta}_2\{(\bar{Z} - \hat{\bar{Z}}_1^o) - h_{1j}(\hat{\bar{Z}}_1^o - z_j)\}.\end{aligned}$$

Thus for $j \in s_2$,

$$\begin{aligned}\hat{\bar{Y}}^c(j) - \hat{\bar{Y}}^c = {} & \{\hat{\bar{Y}}_2^o(j) - \hat{\bar{Y}}_2^o\} - \hat{\beta}_2(j)\{\hat{\bar{X}}_2^o(j) - \hat{\bar{X}}_2^o\} \\ & - \hat{\beta}_1(j)\hat{\beta}_2(j)\{\hat{\bar{Z}}_1^o(j) - \bar{Z}\} \\ & + \hat{\beta}_2[\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_1^o\} - \hat{\beta}_1(j)\{\hat{\bar{Z}}_1^o(j) - \bar{Z}\}] \\ & + \hat{\beta}_2(j)\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_2^o\} \\ & + \hat{\beta}_2[\{\hat{\bar{X}}_2^o(j) - \hat{\bar{X}}_1^o(j)\} \\ & \quad - \hat{\beta}_1(j)\{\bar{Z} - \hat{\bar{Z}}_1^o(j)\} - \hat{\beta}_1\{\bar{Z} - \hat{\bar{Z}}_1^o\}],\end{aligned}$$

and for $j \in (s_1 - s_2)$,

$$\hat{\bar{Y}}^c(j) - \hat{\bar{Y}}^c = \hat{\beta}_2[\{\hat{\bar{X}}_1^o(j) - \hat{\bar{X}}_1^o\} - \hat{\beta}_1(j)\{\hat{\bar{Z}}_1^o(j) - \bar{Z}\}],$$

which proves (2.1).

Consistency of the estimator of variance in (2.2)

In this part of the appendix, we prove that the estimator $\hat{V}_{\text{JACK}}(\hat{\bar{Y}}^c)$ in (2.2) is consistent. First, note that the variance of the estimator $\hat{\bar{Y}}^c$ defined in (1.6) can be approximated as:

$$\begin{aligned}V(\hat{\bar{Y}}^c) \approx {} & V(\hat{\bar{Y}}_2^o) + \beta_2^2[V(\hat{\bar{X}}_1^o) + V(\hat{\bar{X}}_2^o) - 2\text{Cov}(\hat{\bar{X}}_1^o, \hat{\bar{X}}_2^o)] \\ & + \beta_1^2\beta_2^2 V(\hat{\bar{Z}}_1^o) \\ & + 2\beta_2[\text{Cov}(\hat{\bar{Y}}_2^o, \hat{\bar{X}}_1^o) - \text{Cov}(\hat{\bar{Y}}_2^o, \hat{\bar{X}}_2^o)] \\ & - 2\beta_1\beta_2\text{Cov}(\hat{\bar{Y}}_2^o, \hat{\bar{Z}}_1^o) \\ & - 2\beta_1\beta_2^2[\text{Cov}(\hat{\bar{X}}_1^o, \hat{\bar{Z}}_1^o) - \text{Cov}(\hat{\bar{X}}_2^o, \hat{\bar{Z}}_1^o)].\end{aligned}$$

If it is assumed that $\hat{\beta}_1(j) \approx \beta_1$, $\hat{\beta}_2(j) \approx \beta_2$, and similar to Rao and Sitter (1995), that $\bar{x}_n(j)/\bar{x}_r(j) \approx \bar{x}_n/\bar{x}_r$, it is quite straightforward to show that

$$\begin{aligned}\sum_{j\in s}[\hat{\bar{Y}}^c(j)-\hat{\bar{Y}}^c]^2 \approx{}& \sum_{j\in s_2}[\hat{\bar{Y}}_2^o(j)-\hat{\bar{Y}}_2^o]^2+\hat{\beta}_2^2\sum_{j\in s_2}[\hat{\bar{X}}_2^o(j)-\hat{\bar{X}}_2^o]^2\\ &+2\hat{\beta}_2\sum_{j\in s_2}[\hat{\bar{Y}}_2^o(j)-\hat{\bar{Y}}_2^o][\hat{\bar{X}}_1^o(j)-\hat{\bar{X}}_1^o]\\ &-2\hat{\beta}_2\sum_{j\in s_2}[\hat{\bar{Y}}_2^o(j)-\hat{\bar{Y}}_2^o][\hat{\bar{X}}_2^o(j)-\hat{\bar{X}}_2^o]\\ &-2\hat{\beta}_2^2\sum_{j\in s_2}[\hat{\bar{X}}_1^o(j)-\hat{\bar{X}}_1^o][\hat{\bar{X}}_2^o(j)-\hat{\bar{X}}_2^o]\\ &-2\hat{\beta}_1\hat{\beta}_2\sum_{j\in s_2}[\hat{\bar{Y}}_2^o(j)-\hat{\bar{Y}}_2^o][\hat{\bar{Z}}_1^o(j)-\hat{\bar{Z}}_1^o]\\ &-2\hat{\beta}_1\hat{\beta}_2^2\sum_{j\in s_2}[\hat{\bar{X}}_2^o(j)-\hat{\bar{X}}_2^o][\hat{\bar{Z}}_1^o(j)-\hat{\bar{Z}}_1^o]\\ &+\hat{\beta}_2^2\sum_{j\in s}[\hat{\bar{X}}_1^o(j)-\hat{\bar{X}}_1^o]^2\\ &+\hat{\beta}_1^2\sum_{j\in s}[\hat{\bar{Z}}_1^o(j)-\hat{\bar{Z}}_1^o]^2\\ &-2\hat{\beta}_1\hat{\beta}_2^2\sum_{j\in s}[\hat{\bar{X}}_1^o(j)-\hat{\bar{X}}_1^o][\hat{\bar{Z}}_1^o(j)-\hat{\bar{Z}}_1^o].\end{aligned}$$

Since the ten terms on the right hand side of this equation for $\sum_{j\in s}[\hat{\bar{Y}}^c(j)-\hat{\bar{Y}}^c]^2$ are the consistent estimators of the analogous ten terms in the equation above for $V(\hat{\bar{Y}}^c)$, it may be concluded that the jackknife estimator of variance in (2.2) is consistent.

References

Ahmed, M.S. (1997). The general class of chain estimators for the ratio of two means using double sampling. *Communincations in Statistics, Theory and Methods*, 26(9), 2247-2254.

Arnab, R., and Singh, S. (2006). A new method for estimating variance from data imputed with ratio method of imputation. *Statistics and Probability Letters*, 76, 513-519.

Berger, Y. (2007). A jackknife variance estimator for unistage stratified samples with unequal probabilities. *Biometrika*, 94, 953-964.

Berger, Y., and Skinner, C. (2005). A jackknife variance estimator for unequal probability sampling. *Journal of the Royal Statistical Society*, Series B, 67, 79-89.

Chand, L. (1975). *Some ratio type estimators based on two or more auxiliary variables*. PhD Thesis, Iowa State Unversity, Ames, Iowa, USA.

Chen, J., and Shao, J. (2001). Jackknife variance estimation for nearest neighbour imputation. *Journal of the American Statistical Association*, 96, 260-269.

Deville, J.-C., and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

Fuller, W.A. (1998). Replication variance estimation for two-phase samples. *Statistica Sinica*, 8, 117-132.

Hidiroglou, M.A., and Särndal, C.-E. (1995). Use of auxiliary information for two-phase sampling. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Vol. II, 873-878.

Hidiroglou, M.A., and Särndal, C.-E. (1998). Use of auxiliary information for two-phase sampling. *Survey Methodology*, 24, 11-20.

Kim, J.K., Navarro, A. and Fuller, W.A. (2000). Variance estimation for 2000 Census coverage estimates. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 515-520.

Kim, J.K., Navarro, A. and Fuller, W.A. (2006). Replication variance estimation for two-phase stratified sampling. *Journal of the American Statistical Association*, 101, 312-320.

Kim, J.K., and Sitter, R.R. (2003). Efficient replication variance estimation for two-phase sampling. *Statistica Sinica*, 13, 641-653.

Kott, P.S., and Stukel, D. (1997). Can the jackknife be used with a two-phase sample? *Survey Methodology*, 23, 81-89.

Kovar, J., and Chen, E. (1994). Jackknife variance estimation of imputed survey data. *Survey Methodology*, 20, 45-52.

Raj, D. (1965). On sampling over two occasions with probability proportional to size. *Annals of Mathematical Statistics*, 36, 327-330.

Rao, J.N.K., and Sitter, R.R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika*, 82, 453-60.

Singh, S. (2000). Estimation of variance of regression estimator in two phase sampling. *Calcutta Statistical Association Bulletin*, 50, 49-63.

Sitter, R.R. (1997). Variance estimation for the regression estimator in two-phase sampling. *Journal of the American Statistical Association*, 92, 780-787.

Srivenkataramana, T., and Tracy, D.S. (1989). Two-phase sampling for selection with probability proportional to size in sample surveys. *Biometrika*, 76, 818-821.