Survey Methodology, June 1996
Vol. 22, No. 1, pp. 23–31
Statistics Canada

Jackknife Linearization Variance Estimators Under Stratified Multi-Stage Sampling

W. YUNG and J.N.K. RAO[1]

ABSTRACT

Variance estimation for the poststratified estimator and the generalized regression estimator of a total under stratified multi-stage sampling is considered. By linearizing the jackknife variance estimator, a jackknife linearization variance estimator is obtained which is different from the standard linearization variance estimator. This variance estimator is computationally simpler than the jackknife variance estimator and yet leads to values close to the jackknife. Properties of the jackknife linearization variance estimator, the standard linearized variance estimator, and the jackknife variance estimator are studied through a simulation study. All of the variance estimators performed well both unconditionally and conditionally given a measure of how far away the estimated totals of auxiliary variables are from the known population totals. A jackknife variance estimator based on incorrect reweighting performed poorly, indicating the importance of correct reweighting when using the jackknife method.

KEY WORDS: Generalized regression estimator; Jackknife variance estimator; Linearized variance estimator; Poststratified estimator.

1. INTRODUCTION

Large-scale sample surveys often use stratified multi-stage designs with large numbers of strata, L, and relatively few primary sampling units (clusters), $n_h(\geq 2)$, sampled within each stratum. Within each cluster, some elements (ultimate units) are sampled according to some sampling method. We do not specify the number of stages or the sampling methods used after the first-stage sampling, but we assume that subsampling within sampled clusters is performed to ensure unbiased estimation of cluster totals, Y_{hi}, $i = 1, \ldots, n_h$; $h = 1, \ldots, L$.

From the specification of the survey design, basic weights $w_{hik}(> 0)$, attached to the (hik)-th element, are obtained. Often these basic weights w_{hik} are subjected to poststratification adjustment to ensure consistency with known totals of poststratification variables. In the case of a single poststratifier, the weights are ratio-adjusted to the known population counts (*e.g.*, age-sex counts). To handle two or more poststratifiers with known marginal population counts, the weights w_{hik} can be calibrated through generalized regression (see section 4), as in the Canadian Labour Force Survey(CLFS).

The CLFS uses the jackknife method for estimating the variance of the generalized regression estimator. The jackknife method is computer intensive but it is readily applicable to general smooth statistics, unlike the linearization method. Moreover, it possesses good conditional properties. For example, in the context of simple random sampling and the ratio estimator, Royall and Cumberland (1981) showed that the jackknife variance estimator tracks the conditional variance given the sample mean of the auxiliary variable x.

The main purpose of this paper is to study variance estimation for the ratio-adjusted poststratified estimator and the generalized regression estimator under stratified sampling. By linearizing the jackknife variance estimator, a jackknife linearization variance estimator is obtained which is different from the standard linearization variance estimator. In the case of the poststratified estimator, this variance estimator is identical to Rao's (1985) variance estimator. The proposed variance estimator is computationally simpler than the jackknife variance estimator and yet leads to values close to the jackknife.

Section 2 introduces the jackknife variance estimator for the basic expansion estimator of the total, Y. Section 3 presents the jackknife and the jackknife linearization variance estimators for the poststratified estimator. These results are extended in section 4 to the generalized regression estimator in the context of multiple poststratification variables. Section 5 deals with variance estimation for a ratio of two totals, both of which are estimated using a generalized regression estimator. Results of a simulation study on the relative performances of the usual linearization variance estimator, the jackknife and the jackknife linearization variance estimators are reported in section 6.

2. BASIC ESTIMATOR

Using the basic weights w_{hik}, an unbiased estimator of the population total Y is of the form

$$\hat{Y} = \sum_{(hik)\in s} w_{hik} y_{hik}, \tag{2.1}$$

[1] W. Yung, Statistics Canada, Household Survey Methods Division, R.H. Coats Building, Tunney's Pasture, Ottawa, Ontario, K1A 0T6; and J.N.K. Rao, Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario, K1S 5B6.

where s denotes the sample of elements and y_{hik} is the value of the characteristic of interest associated with the sample element $(hik) \epsilon s$. For simplicity, we assume complete response in this paper.

It is common practice to sample clusters without replacement. However, at the stage of variance estimation, the calculations are greatly simplified by treating the sample as if the clusters are sampled with replacement. This approximation generally leads to overestimation of the variance of $\hat{Y}$, but the relative bias is likely to be small if the first-stage sampling fractions are small.

An estimator of the variance of $\hat{Y}$ is given by

$$v(\hat{Y}) = \sum_{h=1}^{L} \frac{1}{n_h(n_h - 1)} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2 = v(y_{hi}), \quad (2.2)$$

where $y_{hi} = \sum_k (n_h w_{hik}) y_{hik}$, and $\bar{y}_h = (1/n_h) \sum_i y_{hi}$. The operator notation $v(y_{hi})$ denotes that $v(\hat{Y})$ depends only on the y_{hi}'s.

To introduce the jackknife method, we need the estimator $\hat{Y}_{(gj)}$ for each (gj) obtained from the sample after omitting the data from the j-th sampled cluster in the g-th stratum ($j = 1, \ldots, n_g$; $g = 1, \ldots, L$). It is simply obtained from (2.1) by letting $w_{gjk} = 0$, changing $w_{gik} (i \neq j)$ to $n_g w_{gik}/(n_g - 1)$ and retaining the original weights w_{hik} for $h \neq g$, *i.e.*,

$$w_{hik(gj)} = \begin{cases} 0 & \text{if} \quad (hi) = (gj) \\ \dfrac{n_g}{(n_g - 1)} w_{gik} & \text{if} \quad h = g \quad \text{and} \quad i \neq j \\ w_{hik} & \text{if} \quad h \neq g. \end{cases}$$

These jackknife weights, $w_{hik(gj)}$, are calculated for each cluster (gj). The resulting estimator of Y is

$$\hat{Y}_{(gj)} = \sum_{(hik) \epsilon s} w_{hik(gj)} y_{hik}.$$

The jackknife variance estimator is then given by

$$v_J(\hat{Y}) = \sum_{g=1}^{L} \frac{n_g - 1}{n_g} \sum_{j=1}^{n_g} (\hat{Y}_{(gj)} - \hat{Y})^2. \quad (2.3)$$

The variance estimator (2.3) is applicable to general smooth statistics, say $\hat{\theta} = g(\hat{Y})$, by simply replacing $\hat{Y}_{(gj)}$ and $\hat{Y}$ with $\hat{\theta}_{(gj)} = g(\hat{Y}_{(gj)})$ and $\hat{\theta}$ respectively. In the linear case, $\hat{\theta} = \hat{Y}$, the jackknife variance estimator is identical to the customary variance estimator (2.2).

3. POSTSTRATIFIED ESTIMATOR

Suppose the population is partitioned into C poststrata with known population counts ${}_cM$, $c = 1, \ldots, C$. We will use the prescript c to denote poststrata. An estimator of ${}_cM$ is given by

$${}_c\hat{M} = \sum_{(hik) \epsilon_c s} w_{hik}, \quad (3.1)$$

where ${}_cs$ is the sample of elements belonging to the c-th poststratum. Similarly, an estimator of the poststratum total ${}_cY$ is

$${}_c\hat{Y} = \sum_{(hik) \epsilon_c s} w_{hik} y_{hik}.$$

Using the estimators ${}_c\hat{Y}$ and ${}_c\hat{M}$, we obtain a poststratified estimator of the total Y as

$$\hat{Y}_{ps} = \sum_c \frac{{}_cM}{{}_c\hat{M}} \, {}_c\hat{Y}. \quad (3.2)$$

We can rewrite (3.2) as

$$\hat{Y}_{ps} = \sum_c \sum_{(hik) \epsilon_c s} {}_cw_{hik} y_{hik}$$

where ${}_cw_{hik} = w_{hik}({}_cM/{}_c\hat{M})$ is the ratio-adjusted weight for $(hik) \epsilon_c s$. If y_{hik} is the indicator variable for a poststratum, say c, then $\hat{Y}_{ps} = {}_cM$, thus ensuring consistency with known totals, ${}_cM$.

The standard linearization variance estimator is given by (2.2) with y_{hi} changed to

$$\tilde{e}_{hi} = \sum_c \sum_{k \epsilon_c s} (n_h w_{hik}) {}_ce_{hik},$$

where ${}_ce_{hik} = y_{hik} - {}_c\hat{Y}/{}_c\hat{M}$ for the k-th element in the (hi)-th cluster belonging to ${}_cs$, *i.e.*,

$$v_L(\hat{Y}_{ps}) = v(\tilde{e}_{hi}). \quad (3.3)$$

Rao (1985) proposed an alternative linearization variance estimator using the ratio-adjusted weights ${}_cw_{hik}$:

$$v_R(\hat{Y}_{ps}) = v(e_{hi}^*) \quad (3.4)$$

where

$$e_{hi}^* = \sum_c \sum_{k \epsilon_c s} (n_h \, {}_cw_{hik}) {}_ce_{hik}.$$

Turning to the jackknife method, we need to recalculate the poststratification weights ${}_cw_{hik}$ each time a cluster (gj) is deleted. This is done by using the jackknife weights $w_{hik(gj)}$ in (3.1) to get ${}_c\hat{M}_{(gj)}$ and then using ${}_cw_{hik(gj)} = ({}_cM/{}_c\hat{M}_{(gj)}) w_{hik(gj)}$ to get

$$\hat{Y}_{ps(gj)} = \sum_{c} \sum_{(hik) \in_c s} {}_c w_{hik(gj)} y_{hik}.$$

The jackknife variance estimator is then obtained as

$$v_J(\hat{Y}_{ps}) = \sum_{g=1}^{L} \frac{n_g - 1}{n_g} \sum_{j=1}^{n_g} (\hat{Y}_{ps(gj)} - \hat{Y}_{ps})^2. \quad (3.5)$$

By linearizing (3.5), we obtain a jackknife linearization variance estimator, $v_{JL}(\hat{Y}_{ps})$, which is identical to Rao's variance estimator (3.4); see also Valliant (1993). In the important special case of $n_h = 2$ clusters per stratum, (3.4) and (3.5) are in fact asymptotically equal to higher order terms, as the number of strata L increases (Yung 1996).

Rao (1985) justified (3.4) on heuristic grounds by noting that for simple random sampling it reduces to a conditionally valid variance estimator given the poststrata sample sizes, unlike the standard linearization variance estimator (3.3). Särndal, Swensson and Wretman (1989) obtained a variance estimator of the form (3.4) in the context of unistage sampling under a model-assisted framework. Since $v_{JL}(\hat{Y}_{ps})$ and $v_J(\hat{Y}_{ps})$ are approximately equal, the foregoing results suggest that both variance estimators should be "robust" in the sense of possessing good conditional properties given the estimated poststrata counts. Valliant (1993) conducted a simulation study to demonstrate the "robustness" of $v_J(\hat{Y}_{ps})$ and $v_{JL}(\hat{Y}_{ps})$.

4. GENERALIZED REGRESSION ESTIMATOR

In practice, it is common to form poststrata according to two or more auxiliary variables. If the resulting cell level population counts are available, the ratio-adjusted poststratified estimator can be used to increase the efficiency of the estimates. However, these cell counts may not be known in practice. For instance, marginal counts may be known only for age groups and race groups but not cell counts for the individual age-race groups. This means that in terms of a two-way table, the marginal counts are known but not the cell level counts. To handle several poststratifiers with known marginal population counts, we can use a generalized regression estimator of Y by using indicator auxiliary variables to denote the categories of the poststratifiers (Huang and Fuller 1978; Deville and Särndal 1992).

Let $\boldsymbol{x}_{hik}$ be a vector of auxiliary variables with known population totals $\boldsymbol{X}$. The generalized regression estimator of Y is then given by

$$\hat{Y}_r = \hat{Y} + (\boldsymbol{X} - \hat{\boldsymbol{X}})^T \hat{\boldsymbol{B}}, \quad (4.1)$$

where

$$\hat{\boldsymbol{X}} = \sum_{(hik) \in s} w_{hik} \boldsymbol{x}_{hik},$$

and $\hat{\boldsymbol{B}}$ is the vector of estimated regression coefficients

$$\hat{\boldsymbol{B}} = \hat{\boldsymbol{A}}^{-1} \hat{\boldsymbol{b}},$$

where

$$\hat{\boldsymbol{A}} = \sum_{(hik) \in s} w_{hik} \boldsymbol{x}_{hik} \boldsymbol{x}_{hik}^T,$$

and

$$\hat{\boldsymbol{b}} = \sum_{(hik) \in s} w_{hik} \boldsymbol{x}_{hik} y_{hik}.$$

The poststratified estimator, $\hat{Y}_{ps}$, is a special case of (4.1) by letting $\boldsymbol{x}_{hik}$ denote the vector of indicator variables for the poststrata. In this case, $\hat{\boldsymbol{X}} = ({}_1\hat{M}, \ldots, {}_C\hat{M})^T$, $\boldsymbol{X} = ({}_1M, \ldots, {}_CM)^T$, and $\hat{\boldsymbol{B}} = ({}_1\hat{R}, \ldots, {}_C\hat{R})^T$ with ${}_c\hat{R} = {}_c\hat{Y}/{}_c\hat{M}$. Thus,

$$\hat{Y}_r = \hat{Y} + \sum_c {}_c\hat{R}({}_cM - {}_c\hat{M}) = \hat{Y}_{ps}.$$

In the case of two or more poststratifiers, $\boldsymbol{X}$ corresponds to the vector of marginal population counts.

The generalized regression estimator may be rewritten as

$$\hat{Y}_r = \sum_{(hik) \in s} w_{hik}^* y_{hik},$$

where

$$w_{hik}^* = w_{hik} a_{hik} \quad (4.2)$$

is the "final" or "calibration" weight with

$$a_{hik} = 1 + \boldsymbol{x}_{hik}^T \hat{\boldsymbol{A}}^{-1} (\boldsymbol{X} - \hat{\boldsymbol{X}}).$$

In the special case of $\hat{Y}_{ps}$, we have $a_{hik} = {}_cM/{}_c\hat{M}$ for $(hik) \in_c s$. Writing $\hat{Y}_r$ in the operator notation as $\hat{Y}_r(y_{hik})$, it is readily verified that the generalized regression estimator $\hat{\boldsymbol{X}}_r = \hat{Y}_r(\boldsymbol{x}_{hik}) = \boldsymbol{X}$, thus ensuring consistency with known totals $\boldsymbol{X}$.

Turning to variance estimation, the standard linearization variance estimator is again given by (2.2) with y_{hi} changed to

$$\tilde{e}_{hi} = \sum_k (n_h w_{hik}) e_{hik},$$

where

$$e_{hik} = y_{hik} - \boldsymbol{x}_{hik}^T \hat{\boldsymbol{B}} \quad (4.3)$$

are the estimated residuals, *i.e.*,

$$v_L(\hat{Y}_r) = v(\tilde{e}_{hi}). \quad (4.4)$$

For the jackknife method we need to recalculate the calibration weights w^*_{hik} each time a cluster (gj) is deleted. These weights are given by

$$w^*_{hik(gj)} = w_{hik(gj)} a_{hik(gj)},$$

where

$$a_{hik(gj)} = 1 + x^T_{hik} \hat{A}^{-1}_{(gj)} (X - \hat{X}_{(gj)}),$$

$$\hat{A}_{(gj)} = \sum_{(hik)\in s} w_{hik(gj)} x_{hik} x^T_{hik},$$

and

$$\hat{X}_{(gj)} = \sum_{(hik)\in s} w_{hik(gj)} x_{hik}.$$

Denote the resulting generalized regression estimator as

$$\hat{Y}_{r(gj)} = \sum_{(hik)\in s} w^*_{hik(gj)} y_{hik}$$

$$= \hat{Y}_{(gj)} + (X - \hat{X}_{(gj)})^T \hat{B}_{(gj)}$$

where $\hat{B}_{(gj)}$ is the vector of estimated regression coefficients when the (gj)-th cluster is deleted:

$$\hat{B}_{(gj)} = \hat{A}^{-1}_{(gj)} \hat{b}_{(gj)}$$

with

$$\hat{b}_{(gj)} = \sum_{(hik)\in s} w_{hik(gj)} x_{hik} y_{hik}.$$

The jackknife variance estimator of $\hat{Y}_r$ is then given by

$$v_J(\hat{Y}_r) = \sum_{g=1}^{L} \frac{n_g - 1}{n_g} \sum_{j=1}^{n_g} (\hat{Y}_{r(gj)} - \hat{Y}_r)^2. \quad (4.5)$$

It is shown in the Appendix that by linearizing the jackknife variance estimator (4.5), one obtains

$$v_{JL}(\hat{Y}_r) = v(e^*_{hi}) \quad (4.6)$$

with

$$e^*_{hi} = \sum_k (n_h w^*_{hik}) e_{hik}$$

where w^*_{hik} is defined in (4.2) and e_{hik} is defined in (4.3). It is interesting to note that the jackknife linearization variance estimator (4.6) is similar to the model-assisted variance estimator proposed by Särndal, Swensson and Wretman (1989) in the context of unistage sampling. Yung (1996) established the asymptotic equivalence of $v_J(\hat{Y}_r)$ and $v_{JL}(\hat{Y}_r)$ to higher order terms in the important special case of $n_h = 2$ clusters per stratum. Note that the above results are also applicable to general auxiliary variables, x_{hik}.

Binder (1996) proposed a new linearization method which also leads to $v_{JL}(\hat{Y}_r)$. In this method, the partial derivatives are evaluated at the estimates $\hat{Y}$, $\hat{X}$ and $\hat{B}$, rather than the population values Y, X and B as in the traditional linearization method. Given that v_J and v_{JL} are design-consistent (Yung 1996) and possess good conditional properties, our results provide theoretical justification for Binder's method which was proposed as a "cookbook approach".

The computation of the jackknife variance estimator involves the inversion of the matrix $\hat{A}_{(gj)}$ for each (gj). However, the jackknife variance estimator can be approximated by retaining the inverse for the full sample, $\hat{A}^{-1}$, and then using modified weights

$$\tilde{w}_{hik(gj)} = w_{hik(gj)} \tilde{a}_{hik(gj)}$$

with

$$\tilde{a}_{hik(gj)} = 1 + (w_{hik}/w_{hik(gj)}) x'_{hik} \hat{A}^{-1} (X - \hat{X}_{(gj)}).$$

The resulting estimator of Y, when the (gj)-th cluster is deleted, is given by

$$\tilde{Y}_{r(gj)} = \sum_{(hik)\in s} \tilde{w}_{hik(gj)} y_{hik}$$

and the corresponding jackknife variance estimator is

$$v_{J1}(\hat{Y}_r) = \sum_{g=1}^{L} \frac{n_g - 1}{n_g} \sum_{j=1}^{n_g} (\tilde{Y}_{r(gj)} - \hat{Y}_r)^2. \quad (4.7)$$

It is readily seen that (4.7) is exactly equal to the standard linearization variance estimator (4.4).

5. ESTIMATION OF A RATIO

Often a ratio of two estimated totals is required. For example, in a family expenditure survey, one may be interested in the proportion of income spent on clothing. Let

$$\hat{Y}_r = \hat{Y} + (X - \hat{X})^T \hat{B}_1$$

be a generalized regression estimator of the total amount spent on clothing, Y. Similarly, let

$$\hat{Z}_r = \hat{Z} + (X - \hat{X})^T \hat{B}_2$$

be a generalized regression estimator of the total income, Z. The proportion of interest is $\theta = Y/Z$, and can be estimated by

$$\hat{\theta} = \hat{Y}_r / \hat{Z}_r.$$

The jackknife variance estimator is given by

$$v_J(\hat{\theta}) = \sum_g \frac{n_g - 1}{n_g} \sum_j (\hat{\theta}_{(gj)} - \hat{\theta})^2 \qquad (5.1)$$

where

$$\hat{\theta}_{(gj)} = \hat{Y}_{r(gj)}/\hat{Z}_{r(gj)}.$$

Linearizing the jackknife variance estimator, (5.1), we obtain a jackknife linearization variance estimator

$$v_{JL}(\hat{\theta}) = v(r_{hi}^{**}) \qquad (5.2)$$

where

$$r_{hi}^{**} = \frac{1}{\hat{Z}_r} \sum_k (n_h w_{hik}^*) e_{hik}^*$$

with

$$e_{hik}^* = e_{hik} - \frac{\hat{Y}_r}{\hat{Z}_r} \tilde{e}_{hik},$$

and

$$e_{hik} = y_{hik} - x_{hik}^T \hat{B}_1, \quad \tilde{e}_{hik} = z_{hik} - x_{hik}^T \hat{B}_2.$$

Proof of (5.2) is omitted for simplicity.

6. SIMULATION STUDY

We performed a simulation study to investigate the unconditional and conditional finite sample properties of the variance estimators in the case of a single poststratifier as well as two poststratification variables. For this purpose, we used a fixed finite population, considered by Valliant (1993), consisting of 10,841 persons included in the September 1988 Current Population Survey (CPS) of the United States. The variable of interest, y, is the weekly wages for each person. The single poststratifier was defined on the basis of age, race and sex, while the two poststratifiers were based on the variables age, with five levels, and race, with two levels (see Tables 1 and 2 for details).

Table 1

Assignment of Age/Race/Sex Categories to Poststrata: Single Poststratifier

Age	Nonblack		Black	
	Male	Female	Male	Female
19 and under	1	1	1	1
20-24	2	3	3	3
25-34	5	6	4	4
35-64	7	8	4	4
65 and over	2	3	3	1

Note: Cell numbers (1-8) are poststratum identification numbers.

Table 2

Assignment of Age/Race Categories to Poststrata: Two Poststratifiers

Age	Nonblack	Black	
19 and under	(1,1)	(1,2)	PS1(1)
20-24	(2,1)	(2,2)	PS1(2)
25-34	(3,1)	(3,2)	PS1(3)
35-64	(4,1)	(4,2)	PS1(4)
65 and over	(5,1)	(5,2)	PS1(5)
	PS2(1)	PS2(2)	

Note: Number in margins are poststratum identification numbers. Cells (i,j) denote poststrata $(i = 1, \ldots, 5; j = 1, 2)$.

The study population contained 2,826 geographical segments, each composed of about four neighbouring households. One hundred design strata ($L = 100$) were created with each stratum having about the same total number of households. We used a stratified two-stage sampling design with segments as clusters and persons as the second-stage units. In each stratum $n_h = 2$ segments were selected with probability proportional to the number of persons in each segment, and a simple random sample of $m_{hi} = 4$ persons was selected without replacement if the sample segment contained more than four persons. In sample segments with four or fewer persons, all persons in the segment were selected. Using this design, we selected two sets of 10,000 independent samples, one set for the one-way poststratification case and the other set for the two-way poststratification case.

From each sample, we computed the basic estimator, the relevant poststratified estimator, $\hat{Y}_{ps}$ or $\hat{Y}_r$, and four variance estimators: the standard linearization variance estimator v_L, the jackknife linearization variance estimator v_{JL}, the jackknife v_J, and an incorrect jackknife variance estimator v_J^*. In applying the jackknife procedure, it is questioned whether or not the "final" or "calibrated" weights need to be recalculated each time a cluster is deleted. The correct jackknife variance estimator does recalculate the "final" weight whenever a cluster is deleted while the incorrect jackknife variance estimator fails to do this. For the one-way poststratification case, $v_J^*(\hat{Y}_{ps})$ uses the full adjustment ${}_cM/{}_c\hat{M}$ instead of ${}_cM/{}_c\hat{M}_{(gj)}$ when the (gj)-th cluster is deleted, *i.e.*, $\hat{Y}_{ps(gj)}$ uses the weights $({}_cM/{}_c\hat{M}) w_{hik(gj)}$ instead of $({}_cM/{}_c\hat{M}_{(gj)}) w_{hik(gj)}$. Similarly, for the two-way poststratification case, $v_J^*(\hat{Y}_r)$ uses the full adjustment a_{hik} instead of $a_{hik(gj)}$ when the (gj)-th cluster is deleted, *i.e.*, $\hat{Y}_r$ uses the weights $w_{hik(gj)} a_{hik}$ instead of $w_{hik(gj)} a_{hik(gj)}$. The linearized version of v_J^* is the same as the variance estimator v_R (equation 3.4) with ${}_ce_{hik}$ replaced by y_{hik} in the case of $\hat{Y}_{ps}$, and v_{JL} (equation 4.6) with e_{hik} replaced by y_{hik} in the case of the generalized regression estimator $\hat{Y}_r$. That is,

$$v_J^*(\hat{Y}_{ps}) = v(y_{hi}^*)$$

with

$$y_{hi}^{*} = \sum_{c} \sum_{k \in_c s} (n_{h\,c} w_{hik}) y_{hik}$$

and

$$v_J^{*}(\hat{Y}_r) = v(y_{hi}^{*})$$

with

$$y_{hi}^{*} = \sum_{k \in s} (n_h w_{hik}^{*}) y_{hik}.$$

Since v_J^* uses the y's instead of the residuals e's, it is clear that v_J^* should overestimate the true variance of the estimator, although it is computationally simpler than v_J.

(i) Unconditional Results

To compare the unconditional performances of the variance estimators we computed the empirical relative bias (RB) for each variance estimator: RB of a variance estimator v is

$$\mathrm{RB} = \frac{1}{\mathrm{MSE}} \left[\frac{1}{10{,}000} \sum_i v_1 \right] - 1$$

where v_i is the value of v for the i-th simulated sample ($i = 1, \ldots, 10{,}000$) and MSE is the empirical MSE of the estimator, say $\tilde{Y}$:

$$\mathrm{MSE} = \frac{1}{10{,}000} \sum_i (\tilde{Y}_i - Y)^2$$

where $\tilde{Y}_i$ is the value of $\tilde{Y}$ in the i-th simulated sample.

Error rates for normal theory confidence intervals on the total Y were also calculated for each variance estimator, using a nominal error rate of 5%:

error rate =

$$1 - \frac{1}{10{,}000} (\text{number of samples with } L_i \le Y \le U_i),$$

where $L_i \le Y \le U_i$ is a confidence interval on Y for the i-th simulated sample. Lower and upper error rates were calculated as:

lower error rate =

$$\frac{1}{10{,}000} (\text{number of samples with } Y < L_i)$$

upper error rate =

$$\frac{1}{10{,}000} (\text{number of samples with } Y > U_i).$$

We also calculated the average lengths of the confidence intervals as

$$\text{average length} = \frac{1}{10{,}000} \sum_i (U_i - L_i).$$

Table 3 reports the unconditional results for the post-stratified estimator $\hat{Y}_{ps}$ using the above performance measures. With respect to relative bias, v_{JL} and v_J both perform well with RB < 1% while the incorrect jackknife v_{JL}^* severely overestimates the MSE (RB = 37%). We note that v_L is also estimating the MSE of $\hat{Y}_{ps}$ well unconditionally (RB < 1%), contrary to Valliant's (1993) claim. Valliant (1993) reported RB of 35% for v_L using the same data set. In view of the design-consistency of v_L supplemented by our simulation results on v_L, we conjecture that Valliant's calculations on v_L might be incorrect.

Table 3

Unconditional Results for the Poststratified Estimator

Performance Measure	$v_L(\hat{Y}_{ps})$	$v_{JL}(\hat{Y}_{ps})$	$v_J(\hat{Y}_{ps})$	$v_J^*(\hat{Y}_{ps})$
Relative bias (%)	−0.44	0.12	0.26	37.16
Error rate (%)	5.20	5.09	5.06	2.41
Lower error rate (%)	2.41	2.35	2.33	0.99
Upper error rate (%)	2.79	2.74	2.73	1.42
Average length	3.81	3.82	3.83	4.48

Turning to confidence interval performance, Table 3 shows that the error rates associated with v_J, v_{JL} and v_L are close to the nominal 5% while the error rate for v_J^* is considerably lower than 5% (about 2.5%). Performances with respect to lower and upper error rates are also similar. The variance estimators, v_J, v_{JL} and v_L, perform similarly in terms of average length of confidence intervals while the average length associated with v_J^* is significantly larger due to overestimation bias. Finally, we note that the performance measures for v_J and v_{JL} are very close, supporting the asymptotic equivalence of v_J and v_{JL}.

Table 4

Unconditional Results for the Generalized Regression Estimator

Performance Measure	$v_L(\hat{Y}_r)$	$v_{JL}(\hat{Y}_r)$	$v_J(\hat{Y}_r)$	$v_J^*(\hat{Y}_r)$
Relative bias (%)	−0.96	0.76	0.57	25.87
Error rate (%)	5.30	5.27	5.23	3.07
Lower error rate (%)	2.24	2.21	2.19	1.08
Upper error rate (%)	3.06	3.06	3.04	1.99
Average length	3.94	3.95	3.95	4.44

Unconditional results for the generalized regression estimator $\hat{Y}_r$ are reported in Table 4. As in the case of $\hat{Y}_{ps}$, the variance estimators v_J, v_{JL} and v_L perform well both in terms of relative bias and error rates of confidence intervals. On the other hand, the incorrect jackknife v_J^* leads to severe overestimation which in turn is reflected in the lower than nominal error rates and larger average length of confidence intervals.

(ii) Conditional Results

We have also studied conditional properties of the variance estimators, following Valliant (1993). For the poststratified estimator, we divided the 10,000 simulated samples into 10 groups each containing 1,000 samples using the measure (Valliant 1993)

$$D_{ps} = \sum_c \left(\frac{{}_c\hat{M}}{{}_cM} - 1 \right).$$

The measure D_{ps} was calculated for each sample and the 10,000 samples were sorted in ascending order according to the D_{ps}-values and then divided into groups. We may interpret D_{ps} as a measure of how "balanced" the sample is with respect to the distribution of the poststrata counts.

For the generalized regression estimator, we used the following natural extension of D_{ps}:

$$D_r = \sum_a \left(\frac{{}_a\hat{M}}{{}_aM} - 1 \right) + \sum_b \left(\frac{{}_b\hat{M}}{{}_bM} - 1 \right),$$

where a and b index the levels of the two poststratification variables and $({}_a\hat{M},\ {}_aM)$ and $({}_b\hat{M},\ {}_bM)$ are the corresponding marginal counts. We may interpret D_r as a measure of how "balanced" the sample is with respect to the distribution of the marginal poststrata counts.

Table 5

Conditional Relative Biases (%) for the Poststratified Estimator

Group	$v_L(\hat{Y}_{ps})$	$v_{JL}(\hat{Y}_{ps})$	$v_J(\hat{Y}_{ps})$	$v_J^*(\hat{Y}_{ps})$
1	−5.00	−8.05	−7.88	17.83
2	0.55	−1.18	−1.01	28.06
3	8.33	7.03	7.19	41.29
4	−1.10	−1.56	−1.42	31.82
5	−0.76	−0.69	−0.55	34.77
6	2.50	3.39	3.53	41.69
7	6.10	7.51	7.66	48.86
8	6.60	8.82	8.96	53.54
9	−4.46	−1.43	−1.31	41.11
10	−13.56	−9.17	−9.07	36.63

Table 6

Conditional Error Rates (%) for the Poststratified Estimator

Group	$v_L(\hat{Y}_{ps})$	$v_{JL}(\hat{Y}_{ps})$	$v_J(\hat{Y}_{ps})$	$v_J^*(\hat{Y}_{ps})$
1	5.5	5.9	5.9	3.4
2	4.6	4.8	4.8	2.9
3	3.7	3.8	3.8	1.9
4	5.7	5.8	5.8	2.9
5	4.9	4.8	4.7	2.6
6	5.1	5.0	4.8	2.2
7	5.2	4.8	4.8	2.1
8	4.5	4.3	4.3	1.3
9	5.8	5.4	5.4	2.4
10	7.0	6.3	6.3	2.4

The results for the poststratified estimator are given in Tables 5 and 6: conditional relative biases in Table 5 and conditional error rates (nominal 5%) in Table 6. These performance measures were computed in the same manner as the unconditional case but from each group separately. It is clear from Tables 5 and 6 that v_J, v_{JL} and v_L all perform well, although v_L is somewhat worse in the extreme groups 1 and 10, while v_J^* performed poorly as before. It is somewhat surprising to see v_L performing so well conditionally. A possible explanation is that with our particular sampling design we have $\hat{M} = \sum_{(hik)\in s} w_{hik} = M$ so that

$$\sum_c {}_c\hat{M} = \hat{M} = M.$$

Because of this, we do not obtain samples which are poorly balanced since if some poststrata counts ${}_c\hat{M}$ are gross overestimates, say, then the other counts correct for the overestimation in order to satisfy the above constraint. Thus, we see mostly well balanced samples in which case v_L is expected to perform well.

Table 7

Conditional Relative Biases (%) for the Generalized Regression Estimator

Group	$v_L(\hat{Y}_r)$	$v_{JL}(\hat{Y}_r)$	$v_J(\hat{Y}_r)$	$v_J^*(\hat{Y}_r)$
1	9.25	4.95	5.13	26.51
2	3.99	1.50	1.67	24.96
3	−3.24	−4.76	−4.59	17.53
4	−2.66	−3.43	−3.26	20.53
5	7.90	7.61	7.80	35.46
6	−3.60	−3.12	−2.94	23.38
7	−9.24	−8.27	−8.08	17.41
8	3.34	5.30	5.50	35.84
9	−3.75	−0.85	−0.62	30.84
10	−8.68	−4.15	−3.92	28.50

Table 8

Conditional Error Rates (%) for the Generalized Regression Estimator

Group	$v_L(\hat{Y}_r)$	$v_{JL}(\hat{Y}_r)$	$v_J(\hat{Y}_r)$	$v_J^*(\hat{Y}_r)$
1	4.3	4.5	4.4	3.0
2	4.9	5.0	5.0	3.3
3	5.0	5.1	5.1	3.8
4	5.7	5.9	5.9	3.3
5	3.9	4.0	4.0	2.3
6	5.7	5.8	5.7	3.0
7	5.9	5.8	5.8	2.9
8	5.8	5.7	5.7	2.8
9	5.5	5.1	4.9	3.0
10	6.3	5.8	5.8	3.3

The results for the generalized regression estimator are given in Tables 7 and 8: conditional relative biases in Table 7 and conditional error rates (nominal 5%) in Table 8. The results are very similar to those for the one stratifier case. In both cases we again note that the performance measures for v_J and v_{JL} are very close, supporting the asymptotic equivalence of v_J and v_{JL}.

In summary, the three variance estimators v_J, v_{JL} and v_L performed similarly. The incorrect jackknife v_J^* performed poorly indicating that reweighting must be done each time a cluster is deleted.

7. CONCLUDING REMARKS

Beebakhee (1995) applied the three variance estimators, v_J, v_{JL} and v_L, to a number of household surveys conducted by Statistics Canada. Her empirical results showed that the jackknife linearization variance estimator, v_{JL}, consistently consumed less time and money for all study surveys than the jackknife variance estimator, v_J, and yet approximated v_J very well. These results are practically important because the users wanted a computationally simpler variance estimator which can approximate the currently used v_J very well. The standard linearization variance estimator v_L performed similar to v_{JL} in terms of cost and time, but it did not approximate v_J as well as v_{JL}.

If the primary interest is the estimation of totals or ratios, then the jackknife linearization variance estimator, v_{JL}, is attractive because it is computationally simpler than the jackknife variance estimator, v_J, and yet leads to values close to the jackknife. But for general smooth statistics v_{JL} suffers from the same disadvantage as the standard linearization variance estimator, v_L, in the sense that both require the derivation of a separate formula for each statistic, unlike v_J. In terms of statistical properties, our simulation study suggests that the three variance estimators, v_J, v_{JL}, and v_L, perform similarly. On the other hand, the incorrect jackknife v_J^*, which uses the same adjustment whenever a cluster is deleted, performs poorly indicating that reweighting must be done each time a cluster is deleted.

ACKNOWLEDGEMENT

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

APPENDIX

Proof of the Result $v_J(\hat{Y}_r) \approx v_{JL}(\hat{Y}_r)$

To establish the desired result, we first approximate the difference $\hat{A}_{(gj)}^{-1} - \hat{A}^{-1}$. Using the matrix identity,

$$(I + PQ)^{-1} = I - P(I + QP)^{-1}Q$$

we get

$$\begin{aligned}\hat{A}_{(gj)}^{-1} - \hat{A}^{-1} &= \hat{A}^{-1}[I + (\hat{A}_{(gj)} - \hat{A})\hat{A}^{-1}]^{-1} - \hat{A}^{-1}\\ &= \hat{A}^{-1}[I - (\hat{A}_{(gj)} - \hat{A})\\ &\qquad (I + \hat{A}^{-1}(\hat{A}_{(gj)} - \hat{A}))^{-1}\hat{A}^{-1}] - \hat{A}^{-1}\\ &\approx -\hat{A}^{-1}(\hat{A}_{(gj)} - \hat{A})\hat{A}^{-1}. \qquad (A.1)\end{aligned}$$

The approximation (A.1) follows by noting that (i) $\hat{A}_{(gj)} - \hat{A}$ is of lower order than $\hat{A}$ under the assumption that no cluster contribution is of disproportionate size as the number of strata L increases (see Yung (1996) for details on regularity conditions) and (ii) $[I + \hat{A}^{-1}(\hat{A}_{(gj)} - \hat{A})]^{-1} \approx I - \hat{A}^{-1}(\hat{A}_{(gj)} - \hat{A})$.

Using (A.1), we obtain

$$\begin{aligned}\hat{B}_{(gj)} - \hat{B} &= (\hat{A}_{(gj)}^{-1} - \hat{A}^{-1} + \hat{A}^{-1})(\hat{b}_{(gj)} - \hat{b} + \hat{b})\\ &\qquad - \hat{A}^{-1}\hat{b}\\ &\approx (\hat{A}_{(gj)}^{-1} - \hat{A}^{-1})\hat{b} + \hat{A}^{-1}(\hat{b}_{(gj)} - \hat{b})\\ &\approx -\hat{A}^{-1}(\hat{A}_{(gj)} - \hat{A})\hat{B} + \hat{A}^{-1}(\hat{b}_{(gj)} - \hat{b}). \qquad (A.2)\end{aligned}$$

It now follows from (A.2) that

$$\begin{aligned}\hat{Y}_{r(gj)} - \hat{Y}_r &\approx (\hat{Y}_{(gj)} - \hat{Y}) - (\hat{X}_{(gj)} - \hat{X})^T\hat{B}\\ &\qquad - (\hat{X} - X)^T(\hat{B}_{(gj)} - \hat{B})\\ &\approx \frac{1}{n_g - 1}(\bar{e}_g^* - e_{gj}^*), \qquad (A.3)\end{aligned}$$

where $e^*_{gj} = \sum_k (n_g w^*_{gjk}) e_{gjk}$ and $\bar{e}^*_g = (1/n_g) \sum_j e^*_{gj}$. We used the following results in arriving at (A.3):

$$(\hat{Y}_{(gj)} - \hat{Y}) - (\hat{X}_{(gj)} - \hat{X})^T \hat{B} = \frac{1}{n_g - 1} (\bar{e}_g - e_{gj})$$

and

$$(\hat{X} - X)^T (\hat{B}_{(gj)} - \hat{B}) \approx$$

$$(X - \hat{X})^T \hat{A}^{-1} \left[\frac{1}{n_g - 1} (\bar{u}_g - u_{gj}) \right],$$

where $e_{gj} = \sum_k (n_g w_{gjk}) e_{gjk}$ and $u_{gj} = \sum_k (n_g w_{gjk}) x_{gjk} e_{gjk}$.

It now follows from (A.3) that

$$v_J(\hat{Y}_r) \approx \sum_{h=1}^{L} \frac{1}{n_h(n_h - 1)} \sum_{i=1}^{n_h} (e^*_{hi} - \bar{e}^*_h)^2$$

$$= v(e^*_{hi}) = v_{JL}(\hat{Y}_r).$$

REFERENCES

BEEBAKHEE, R. (1995). A comparison of two variance estimation methods: The Jackknife and the linearized Jackknife. Methodology Branch Working Paper, HSMD-95-005E. Statistics Canada.

BINDER, D.A. (1996). Linearization methods for single phase and two phase samples: A cookbook approach. *Survey Methodology*, 22, 17-22.

CASADY, R.J., and VALLIANT, R. (1993). Conditional properties of poststratified estimators under normal theory. *Survey Methodology*, 19, 183-192.

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

HUANG, E.T., and FULLER, W.A. (1978). Nonnegative regression estimation for sample survey data. *Proceedings of the Social Statistics Section, American Statistical Association*, 300-305.

RAO, J.N.K. (1985). Conditional inference in survey sampling. *Survey Methodology*, 11, 15-31.

ROYALL, R.M., and CUMBERLAND, W.G. (1981). An empirical study of the ratio estimator and estimator of its variance. *Journal of the American Statistical Association*, 76, 66-88.

SÄRNDAL, C.E., SWENSSON, B., and WRETMAN, J. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika*, 76, 527-537.

STATISTICS CANADA (1990). *Methodology of the Canadian Labour Force Survey*. Catalogue No. 71-526.

VALLIANT, R. (1993). Poststratification and conditional variance estimation. *Journal of the American Statistical Association*, 88, 89-96.

YUNG, W. (1996). Contributions to poststratification in stratified multi-stage samples. Unpublished Ph.D. thesis, Carleton University, Ottawa, Canada.