Survey Methodology, June 1996
Vol. 22, No. 1, pp. 17–22
Statistics Canada

Linearization Methods for Single Phase and Two-Phase Samples: A Cookbook Approach

DAVID A. BINDER[1]

ABSTRACT

There are a number of asymptotically equivalent procedures for deriving the Taylor series approximation of variances for complex statistics. In Binder and Patak (1994) the theoretical justification for one class of methods was derived. However, many of these methods can be derived for practical examples using straightforward techniques that are not clearly described in Binder and Patak. In this paper we give a "cookbook" approach that can be used for many examples, and that has been shown to have good finite sample properties. Normally the method of choice becomes clear through arguments such as model-assisted methods or linearizing the jackknife; however, using our approach yields the desired results more directly. As well, we present new results on the application of these techniques to two-phase samples.

KEY WORDS: Complex surveys; Variance estimation; Ratio estimator; Regression estimator; Wilcoxon rank sum test; Estimating equations.

1. THE METHOD

The derivation of the asymptotic variance for a wide class of estimators from complex survey samples is now well established in the literature, at least to a first order approximation. However, there are a number of competing estimators of the variance, all of which are asymptotically equivalent. In this paper, we discuss a simple derivation of one of the most favoured of these estimators in a general setting. This simple derivation is useful for practitioners, who may be baffled by the choices available, and need a quick solution to the problem.

We start with a simple example of the approach using the ratio estimator of a population total. Here the estimator is

$$\hat{Y}_R = \hat{R}X, \tag{1}$$

for

$$\hat{R} = \hat{Y}/\hat{X}, \quad \text{and} \quad \hat{Y} = \sum_{k \in s} w_k y_k,$$

where, s is the set of indices corresponding to sampled units and w_k is the sampling weight, normalized so that $\sum w_k$ is an estimator of the population total; *e.g.*, $w_k = 1/\pi_k$, where π_k is the first order inclusion probability. The definition of $\hat{X}$ is analogous to that of $\hat{Y}$. Applying total differentials to both sides of (1), we obtain

$$(d\hat{Y}_R) = (d\hat{R})X, \tag{2a}$$

where

$$(d\hat{R}) = \frac{(d\hat{Y})}{\hat{X}} - \frac{\hat{Y}}{\hat{X}^2}(d\hat{X}) \tag{2b}$$

$$= \frac{1}{\hat{X}}[(d\hat{Y}) - \hat{R}(d\hat{X})].$$

We note that, in general, the total differential for $\hat{T} = g(\hat{Y}_1, \ldots, \hat{Y}_m)$ is given by

$$(d\hat{T}) = \sum \left[\frac{\partial g(\hat{Y})}{\partial \hat{Y}_i}\right](d\hat{Y}_i).$$

Although we could have avoided using $\hat{R}$ in (1) by simply defining

$$Y_{\hat{R}} = \frac{\hat{Y}}{\hat{X}}X,$$

thus removing the need for explicitly defining $(d\hat{R})$ in (2b), we did so to make the more complex examples, to be given in Section 1.2, clearer. We also note that (2a) does not include the total differential of X, the population total of the x-variable, since X is assumed to be fixed and known.

The next step is to replace all total differentials of estimated quantities by deviations from the their respective expected values. On the right hand side, we substitute for $(d\hat{Y})$ the expression $(\sum w_k y_k - Y)$, and so on. For the quantity of interest, $\hat{Y}_R$, we replace $d\hat{Y}_R$ by $\hat{Y}_R - Y$. From (2), performing this step, yields

$$\hat{Y}_R - Y \doteq \frac{X}{\hat{X}}\left[\left(\sum w_k y_k - Y\right) - \hat{R}\left(\sum w_k x_k - X\right)\right]. \tag{3}$$

[1] David A. Binder, Director, Business Survey Methods Division, Statistics Canada, R.H. Coats Building, 11 "A", Ottawa, Ontario, Canada, K1A 0T6.

We see that this expression contains a number of weighted estimators - those that explicitly show their dependence on the w_k's, ($\sum w_k y_k$ and $\sum w_k x_k$) and those where the w_k's are implicit in the expression ($\hat{X}$ and $\hat{R}$).

For the last step, we isolate z_k, defined by rewriting (3) as

$$\hat{Y}_R - Y \doteq \sum w_k z_k + \text{other terms not depending } \textit{explicitly} \text{ on } w_k.$$

Here, we obtain

$$z_k = \frac{X}{\hat{X}}(y_k - \hat{R}x_k). \tag{4}$$

The justification for ignoring the terms not depending explicitly on w_k will be given in Section 4. Note that $\sum w_k z_k$ has the form of the estimate of the population total of the variable z.

Now to obtain the variance of $\hat{Y}_R$, we insert the new variable z_k into the k-th sample record, and use a standard procedure for estimating the variance of a total, applied to this variable. It is assumed that a variance estimator with good properties is available for the sample design under consideration.

A summary of the method in general is the following:

1. We let the estimator of T be $\hat{T}$ and take its total differential. We assume that $\hat{T}$ is asymptotically design consistent.
2. We replace total differential of $\hat{T}$, $d\hat{T}$, by $\hat{T} - T$. We replace all other total differentials of estimated quantities by the deviation from their respective expected values, where we substitute for $(d\hat{Y})$ the expression $(\sum w_k y_k - Y)$, and so on.
3. The last step is to isolate z_k, when we rewrite the result of Step 2 as

$$\hat{T} - T \doteq \sum w_k z_k + \text{other terms not depending } \textit{explicitly} \text{ on } w_k.$$

4. Finally, to obtain the estimated variance of $\hat{T}$, we insert the new variable z_k into each sampled record, and use the standard procedure (known to have good properties) for estimating the variance of a total, applied to this variable.

1.1 Simplest General Case

For one-phase samples, a simple general case is where the estimator can be expressed as a differentiable function of the estimated totals for certain survey variables, some of which may be derived variables at the final sampling unit level. In this case our approach gives:

$$\hat{T} = g(\hat{Y}_1, \ldots, \hat{Y}_m)$$

$$(d\hat{T}) = \sum \left[\frac{\partial g(\hat{Y})}{\partial \hat{Y}_i}\right](d\hat{Y}_i)$$

$$\begin{aligned} \hat{T} - T &\doteq \sum_i \left[\frac{\partial g(\hat{Y})}{\partial \hat{Y}_i}\right]\left(\sum_k w_k y_{ik} - Y_i\right) \\ &= \sum w_k z_k + \ldots, \end{aligned} \tag{5}$$

where

$$z_k = \sum_i \left[\frac{\partial g(\hat{Y})}{\partial \hat{Y}_i}\right] y_{ik} = \left[\frac{\partial g(\hat{Y})}{\partial \hat{Y}}\right]' y_k. \tag{6}$$

In what way is this formulation different from standard Taylor methods? The main difference is how expression (5) is treated. In standard methods, the partial derivatives are evaluated at their expected values before z_k is derived. Then, for those components of z_k that are unknown, an estimator is substituted. For the ratio estimator, (1), this would result in $X/\hat{X}$ disappearing from z_k in (4), since when $\hat{X}$ is replaced by its expected value, $X/\hat{X}$ becomes unity. The $\hat{R}$ remains in the expression, as it is used to estimate R, which is needed in the usual derivation of z_k.

Kott (1990) argues that the variance estimator for the ratio which we have derived has good conditional properties compared to the estimator which leaves out the factor $X/\hat{X}$. A number of others have come to similar conclusions. Rao (1995) showed that the method agrees with that obtained from the linearized jackknife. Our conjecture is that since the partial derivatives in expression (5) are evaluated at $\hat{Y}$ rather than Y, the linearization is "closer" to the original statistic, $\hat{T}$, so that the resulting variances have better properties. This is, of course, not a technical statement, but rather an intuitive justification of the method.

We note that in expression (6) for z_k, all the terms are directly observed from the sample, so that no substitution of estimators for unknown quantities is needed.

1.2 The Case with Extra Parameters

For many examples, the estimator is most easily defined in terms that include the use of parameters that are only used to simplify the definition of the parameter of interest. For the ratio estimator, $\hat{R}$ is an example of such an *extra parameter*. In this case, an explicit equation for the estimator of the extra parameter is available. The general method in the presence of extra parameters may be written as:

$$\hat{T} = g_1(\hat{Y}_1, \ldots, \hat{Y}_m, \hat{\lambda}), \quad \text{where } \hat{\lambda} = g_2(\hat{Y}_1, \ldots, \hat{Y}_m),$$

$$(d\hat{T}) = \sum \left[\frac{\partial g_1(\hat{Y}, \hat{\lambda})}{\partial \hat{Y}_i}\right](d\hat{Y}_i) + \sum \left[\frac{\partial g_1(\hat{Y}, \hat{\lambda})}{\partial \hat{\lambda}_j}\right](d\hat{\lambda}_j),$$

where

$$(d\hat{\lambda}_j) = \sum_i \left[\frac{\partial g_{2j}(\hat{Y})}{\partial \hat{Y}_i}\right](d\hat{Y}_i),$$

$$\hat{T} - T \doteq \sum \frac{\partial g_1(\hat{Y},\hat{\lambda})}{\partial \hat{Y}_i}\left(\sum_k w_k y_{ik} - Y_i\right)$$

$$+ \sum \frac{\partial g_1(\hat{Y},\hat{\lambda})}{\partial \hat{\lambda}_j} \sum_i \frac{\partial g_{2j}(\hat{Y})}{\partial \hat{Y}_i}\left(\sum_k w_k y_{ik} - Y_i\right)$$

$$= \sum w_k z_k + \ldots,$$

where

$$z_k = \left[\frac{\partial g_1(\hat{Y},\hat{\lambda})}{\partial \hat{Y}}\right]' y_k + \left[\frac{\partial g_1(\hat{Y},\hat{\lambda})}{\partial \hat{\lambda}}\right]'\left[\frac{\partial g_2(\hat{Y})}{\partial \hat{Y}}\right] y_k. \quad (7)$$

For the case where the extra parameters are defined only implicitly through estimating equations, we have the following generalization:

$$\hat{T} = g(\hat{Y}_1, \ldots, \hat{Y}_m, \hat{\lambda}),$$

where

$$\hat{U}(\hat{Y}_1, \ldots, \hat{Y}_m, \hat{\lambda}) = 0. \quad (8)$$

$$(d\hat{T}) = \sum \left[\frac{\partial g(\hat{Y},\hat{\lambda})}{\partial \hat{Y}_i}\right](d\hat{Y}_i) + \left[\frac{\partial g(\hat{Y},\hat{\lambda})}{\partial \hat{\lambda}}\right]'(d\hat{\lambda}),$$

where by taking the total differential of (8) and isolating $(d\hat{\lambda})$, we have

$$(d\hat{\lambda}) = -\left[\frac{\partial \hat{U}(\hat{Y},\hat{\lambda})}{\partial \hat{\lambda}}\right]^{-1} \sum \left[\frac{\partial \hat{U}(\hat{Y},\hat{\lambda})}{\partial \hat{Y}_i}\right](d\hat{Y}_i). \quad (9)$$

$$\hat{T} - T \doteq \sum_i \left(\frac{\partial g}{\partial \hat{Y}_i}\right)\left(\sum_k w_k y_{ik} - Y_i\right)$$

$$- \left(\frac{\partial g}{\partial \hat{\lambda}}\right)'\left[\frac{\partial \hat{U}}{\partial \hat{\lambda}}\right]^{-1} \sum_i \left(\frac{\partial \hat{U}}{\partial \hat{Y}_i}\right)\left(\sum_k w_k y_{ik} - Y_i\right)$$

$$= \sum w_k z_k + \ldots,$$

where

$$z_k = \left[\frac{\partial g}{\partial \hat{Y}}\right]' y_k - \left[\frac{\partial g}{\partial \hat{\lambda}}\right]'\left[\frac{\partial \hat{U}}{\partial \hat{\lambda}}\right]^{-1}\left[\frac{\partial \hat{U}}{\partial \hat{Y}}\right]' y_k. \quad (10)$$

We see, of course, that (10) is a generalization of the previous forms for z_k given in (6) and (7).

2. OTHER EXAMPLES

Expressions (6), (7) and (10) above are displayed only for the purpose of giving the specific formulae for the various cases. However, in practice, we recommend using the basic steps from first principles. To demonstrate this, we give two examples: one is the familiar Generalized Regression Estimator (GREG); the other gives some new results for the Wilcoxon Rank Sum Test statistic for data from complex surveys.

2.1 Generalized Regression Estimator

The usual Generalized Regression Estimator, given, for example, in Särndal, Swensson and Wretman (1989), may be written as

$$\hat{Y}_{GREG} = \hat{Y} + \hat{\beta}'(X - \hat{X}), \quad (11)$$

where the extra parameter $\hat{\beta}$ is defined as the solution to

$$\sum_k w_k x_k (y_k - x_k' \hat{\beta})/c_k = 0,$$

where c_k is the factor to allow for heteroscedastic variance in the regression model. This is equivalent to

$$\hat{S}_{xx}\hat{\beta} - \hat{S}_{xy} = 0, \quad (12)$$

with obvious definitions for $\hat{S}_{xx}$ and $\hat{S}_{xy}$. Taking total differentials in (12) we get

$$(d\hat{S}_{xx})\hat{\beta} + \hat{S}_{xx}(d\hat{\beta}) - (d\hat{S}_{xy}) = 0,$$

so that

$$(d\hat{\beta}) = \hat{S}_{xx}^{-1}[(d\hat{S}_{xy}) - (d\hat{S}_{xx})\hat{\beta}].$$

Therefore, we have

$$\hat{\beta} - \beta \doteq \sum w_k \hat{S}_{xx}^{-1}[x_k(y_k - x_k'\hat{\beta})]/c_k + \ldots .$$

Now, taking total differentials of (11), we have

$$(d\hat{Y}_{GREG}) = (d\hat{Y}) - \hat{\beta}'(d\hat{X}) + (d\hat{\beta})'(X - \hat{X})$$

$$= (d\hat{Y}) - \hat{\beta}'(d\hat{X}) + [(d\hat{S}_{xy}') - \hat{\beta}'(d\hat{S}_{xx})]\hat{S}_{xx}^{-1}(X - \hat{X}).$$

After some algebraic manipulation, we obtain

$$\hat{Y}_{GREG} - Y = \sum w_k e_k [1 + x_k' \hat{S}_{xx}^{-1}(X - \hat{X})/c_k] + \ldots,$$

where $e_k = y_k - x_k'\hat{\beta}$. We, therefore, define

$$z_k = e_k[1 + x_k'\hat{S}_{xx}^{-1}(X - \hat{X})/c_k].$$

Taking the variance of the estimated total of this z-variable is identical to the variance proposed in Särndal, Swensson and Wretman (1989). There, it is argued on the basis of the validity of the regression model, that this variance is preferred to other Taylor expansion estimators for the variance. We see that the derivation of this z-variable is natural in our approach.

2.2 Wilcoxon Rank Sum Statistic

We now show how our method works in the case of a more difficult non-standard case. We assume that our sampled units belong to one of two subpopulations which we name Population 1 and Population 2. We define

$$I\{x \le y\} = \begin{cases} 1 \text{ if } x \le y, \\ 0 \text{ otherwise,} \end{cases} \text{ and } \delta_k = \begin{cases} 1 \text{ if } k \in \text{Pop. } 1 \\ 0 \text{ otherwise.} \end{cases}$$

We let

$$\hat{N}_1(t) = \sum_{k \in s} w_k \delta_k I\{x_k \le t\},$$

which corresponds to the estimated number of Population 1 units that have values less than or equal to t. We define $\hat{N}_2(t)$ analogously. We denote $\hat{N}_j = \hat{N}_j(\infty)$, the estimated number of units in Population j. Now a weighted version of the Wilcoxon Rank Sum Test statistic is

$$\hat{T}_W = \int_0^\infty [\hat{N}_1(t) + \hat{N}_2(t)]d\hat{N}_1(t). \tag{13}$$

This corresponds to the weighted sum of the ranks from Population 1 among the weighted ranks of the combined sample. To derive the asymptotic expected value of $\hat{T}_W$ in (13), we let $N_i(t) = E[\hat{N}_i(t)]$ for $i = 1, 2$, and substitute $N_i(t)$ for $\hat{N}_i(t)$ in (13). We then define $F_i(t) = N_i(t)/N_i$, where $N_i = E(\hat{N}_i)$ and we give the null hypothesis as $F_1(t) = F_2(t) = F(t)$, say. This results in the asymptotic expectation being

$$\int_0^1 (N_1 + N_2)F(t)N_1 dF(t) = N_1(N_1 + N_2)/2.$$

Note that in the case of independent samples of size N_1 and N_2 from Population 1 and Population 2, respectively, where each population is assumed to have a continuous distribution function and the samples are taken using simple random sampling, the exact expected value for $\hat{T}_W$ in (13) is $N_1(N_1 + N_2 + 1)/2$.

We consider the statistic

$$\hat{T}_W^* = \int_0^\infty [\hat{N}_1(t) + \hat{N}_2(t)]d\hat{N}_1(t) - \frac{\hat{N}_1(\hat{N}_1 + \hat{N}_2)}{2}.$$

We use Δ rather than d to denote the total differential, since d is used under the integral. Therefore, we have

$$\begin{aligned}(\Delta\hat{T}_W^*) = &\int_0^\infty [\Delta\hat{N}_1(t) + \Delta\hat{N}_2(t)]d\hat{N}_1(t) \\ &+ \int_0^\infty [\hat{N}_1(t) + \hat{N}_2(t)]d\Delta\hat{N}_1(t) \\ &- \frac{(\Delta\hat{N}_1)(\hat{N}_1 + \hat{N}_2) + \hat{N}_1(\Delta\hat{N}_1 + \Delta\hat{N}_2)}{2}.\end{aligned}$$

Continuing with our usual approach, we have

$$\begin{aligned}\hat{T}_W^* - T_W^* \doteq &\int_0^\infty \left(\sum w_k I\{x_k \le t\}\right) d\hat{N}_1(t) \\ &+ \sum w_k \delta_k [\hat{N}_1(x_k) + \hat{N}_2(x_k)] \\ &- \frac{\sum w_k \delta_k (\hat{N}_1 + \hat{N}_2) + \hat{N}_1 \sum w_k}{2} + \ldots,\end{aligned}$$

so that

$$\begin{aligned}z_k = &\sum_j w_j \delta_j I\{x_k \le x_j\} + \delta_k[\hat{N}_1(x_k) + \hat{N}_2(x_k)] \\ &- \frac{\delta_k(\hat{N}_1 + \hat{N}_2) + \hat{N}_1}{2}.\end{aligned} \tag{14}$$

We are not aware of this result previously being documented. It can be shown that when the null hypothesis is true and we select independently from two populations using simple random sampling, where the populations have continuous distribution functions, the variance we obtain from the z-variables in (14) is asymptotically equivalent to the usual classical formula.

3. TWO-PHASE SAMPLES

The method described above extends quite easily to the case of two-phase samples. For example, consider the two-phase ratio estimator of the population total, given by

$$\hat{Y}_{R(2)} = \frac{\hat{Y}}{\hat{X}}\hat{X}^{(1)} = \hat{R}\hat{X}^{(1)}, \tag{15}$$

where $\hat{X}^{(1)} = \sum w_k x_k$ is the first phase estimate of X based on first phase weights $\{w_k\}$, and $\hat{Y}$ and $\hat{X}$ are the estimates of Y and X, respectively, based the second phase sample units with weights $\{w_k w_{2k}\}$, where w_{2k} is the weight assigned to the selected second phase unit, conditional on being in the first phase sample. In particular, letting

$$a_k = \begin{cases} 1 & \text{if the } k\text{-th unit is in the second phase sample,} \\ 0 & \text{otherwise,} \end{cases}$$

we have

$$\hat{Y} = \sum_{k \in s} w_k w_{2k} a_k y_k,$$

where s is the set of indices corresponding to units in the first phase sample.

Taking total differentials of (15), we have

$$(d\hat{Y}_{R(2)}) = \left(\frac{\hat{X}^{(1)}}{\hat{X}}\right) [(d\hat{Y}) - \hat{R}(d\hat{X})] + \hat{R}(d\hat{X}^{(1)}).$$

We now replace the total differentials by weighted sums over first phase units:

$$\hat{Y}_{R(2)} - \hat{Y} \doteq$$

$$\sum_{k \in s} w_k \left[a_k w_{2k} \left(\frac{\hat{X}^{(1)}}{\hat{X}}\right) (y_k - \hat{R}x_k) + \hat{R}x_k \right] + \ldots,$$

so that

$$z_k = a_k w_{2k} \left(\frac{\hat{X}^{(1)}}{\hat{X}}\right) (y_k - \hat{R}x_k) + \hat{R}x_k. \qquad (16)$$

We see that the steps we have taken are essentially the same as in the one phase sample case. However, it is important to note that now z_k contains the random variable, a_k, that is used to indicate whether or not the sample unit is in the second phase sample. This is needed to compute the two phase variance estimator.

Variances obtained from the z-variable in (16) are identical to those given in Rao and Sitter (1995), who used a linearization of the jackknife to obtain their results.

Extensions to other estimation problems in two phase samples are straightforward. Suppose, for example, that $(\hat{Y}_1, \ldots, \hat{Y}_m)$ are estimates of $(Y_1, \ldots, Y_m)$ from the second phase samples, and that $(\hat{X}_1^{(1)}, \ldots, \hat{X}_p^{(1)})$ are estimates of variables available only for first phase sample units. We suppose that a set of extra parameters, λ, are defined only in terms of the units in the second phase, and that the variable of interest is defined in terms of these extra parameters and the $\hat{X}_j^{(1)}$'s. Formally, then, we have

$$U(\hat{\lambda}, \hat{Y}) = 0,$$

and

$$\hat{T} = g(\hat{X}^{(1)}, \hat{\lambda}).$$

Taking total differentials, we have as in (9),

$$(d\hat{\lambda}) = -\left[\frac{\partial \hat{U}}{\partial \hat{\lambda}}\right]^{-1} \left[\frac{\partial \hat{U}}{\partial \hat{Y}}\right] (d\hat{Y}),$$

so that

$$\hat{T} - T \doteq \left[\frac{\partial g}{\partial \hat{X}^{(1)}}\right]' \left(\sum_k w_k x_k - X\right)$$

$$- \left[\frac{\partial g}{\partial \hat{\lambda}}\right]' \left[\frac{\partial \hat{U}}{\partial \hat{\lambda}}\right]^{-1} \left[\frac{\partial \hat{U}}{\partial \hat{Y}}\right] \left(\sum_k a_k w_k w_{2k} y_k - Y\right).$$

Therefore, the general expression for z_k is

$$z_k = \left[\frac{\partial g}{\partial \hat{X}^{(1)}}\right]' x_k - \left[\frac{\partial g}{\partial \hat{\lambda}}\right]' \left[\frac{\partial \hat{U}}{\partial \hat{\lambda}}\right]^{-1} \left[\frac{\partial \hat{U}}{\partial \hat{Y}}\right] a_k w_{2k} y_k.$$

It then becomes necessary to put the z-variable into the algorithm that estimates the variance of the estimator of a total from a two phase sample.

4. JUSTIFICATION

The technique we have described can be considered as a direct result of the formulation given in Binder and Patak (1994). We will summarize one of the main results in that paper. Suppose we are interested in parameter θ, defined as the solution to

$$\hat{U}_1(\theta, \hat{\lambda}_\theta) = \sum_{k \in s} w_k u_1(y_k, \theta, \hat{\lambda}_\theta) = 0,$$

where $\hat{\lambda}_\theta$ is the estimate of an extra parameter, defined as the solution to

$$\hat{U}_2(\theta, \hat{\lambda}_\theta) = \sum_{k \in s} w_k u_2(y_k, \theta, \hat{\lambda}_\theta) = 0,$$

for a given θ. Through an argument based on removing extra parameters for problems of testing hypotheses on θ, Binder and Patak recommend basing inferences about θ on the variable

$$u^* = u_1(y, \theta, \hat{\lambda}_\theta) - \left[\frac{\partial \hat{U}_1}{\partial \hat{\lambda}_\theta}\right] \left[\frac{\partial \hat{U}_2}{\partial \hat{\lambda}_\theta}\right]^{-1} u_2(y, \theta, \hat{\lambda}_\theta). \qquad (17)$$

In particular, two-sided confidence intervals for θ are to be based on

$$\left\{\theta \,\middle|\, \frac{\hat{U}_1^2(\theta, \hat{\lambda}_\theta)}{\hat{W}} \leq \chi^2_{1-\alpha}(1)\right\},$$

where $\hat{W}$ is the estimated variance of the estimator of a total when the variable being estimated is u^*.

We let $u_1 = g(\lambda_1, \lambda_2) - \theta$. The kernel of the estimating equations for the y-totals will be given by $u_{21} = y - \lambda_1$ and the kernel of the estimating equations for λ_2 is given by $u_{22}(\lambda_1, \lambda_2)$. We let

$$\hat{U}_2 = \sum w_k \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix} = \begin{bmatrix} \hat{Y} - \hat{N}\hat{\lambda}_1 \\ \hat{N}u_{22} \end{bmatrix}, \text{ where } \hat{N} = \sum w_k.$$

After some algebra, from (17) the variance of interest is the variance of the estimated total based on the variable u^*, given by,

$$\left[\frac{\partial g(\hat{\lambda}_1, \hat{\lambda}_2)}{\partial \hat{\lambda}_1}\right]' y - \left[\frac{\partial g(\hat{\lambda}_1, \hat{\lambda}_2)}{\partial \hat{\lambda}_2}\right]' \left[\frac{\partial u_{22}(\hat{\lambda}_1, \hat{\lambda}_2)}{\partial \hat{\lambda}_2}\right]^{-1} \left[\frac{\partial u_{22}(\hat{\lambda}_1, \hat{\lambda}_2)}{\partial \hat{\lambda}_1}\right] y + \text{constant terms}.$$

This is equivalent to expression (10), thus showing that the methods here are consistent with those in Binder and Patak (1994).

ACKNOWLEDGEMENTS

I wish to thank Georgia Roberts and Alain Théberge who had many useful suggestions on improving the readability of an earlier draft. I am also grateful to J.N.K. Rao for many useful discussions on this topic, and to an anonymous referee for some constructive remarks.

REFERENCES

BINDER, D.A., and PATAK, Z. (1994). Use of estimating functions for interval estimation from complex surveys. *Journal of the American Statistical Association*, 89, 1035-1043.

KOTT, P.S. (1990). Estimating the conditional variance of a design consistent regression estimator. *Journal of Statistical Planning and Inference*, 24, 287-296.

RAO, J.N.K. (1995). Private communication.

RAO, J.N.K., and SITTER, R.R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika*, 82, 453-460.

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