SSC Annual Meeting, June 1999
Proceedings of the Survey Methods Section

VARIANCE COMPUTATION FOR COMPLEX SURVEYS USING ESTIMATING EQUATIONS

M.A. Hidiroglou, J.N.K. Rao and W. Yung[1]

ABSTRACT

Linear or nonlinear parameters of interest, such as population means, ratios, linear and logistic regression coefficients can be expressed as solutions to suitable census "estimating equations" (Binder, 1983). Parameter estimates are obtained by solving sample estimating equations which involve the design weights as well as adjustment factors (also called g-weights) based on auxiliary information. Using Binder's (1983) Taylor linearization method, we obtain standard errors of parameter estimates involving g-weights such as those resulting from post-stratification or regression adjustments. These standard errors incorporate the g-weights as well as synthetic residuals obtained by regressing the components of the estimating functions on the auxiliary variables. For stratified simple random sampling, we also obtain standard errors by linearizing the jackknife variance estimator. We show how this methodology could be implemented in a computer system to provide a unified, flexible approach of calculating estimates and associated errors.

KEY WORDS: Estimating equations; Jackknife linearization; Taylor linearization; Variance estimation

RÉSUMÉ

Les paramètres linéaires ou non linéaires d'intérêt, tels que les moyennes d'une population, les proportions, les coefficients de régression logistique et linéaire peuvent être exprimés sous forme de solutions aux " équations estimantes" de la population (Binder, 1983, Thompson, 1997). Les estimations de paramètre sont obtenues en résolvant des équations estimantes basées sur l'échantillon. Ces équations sont en fonction des poids de sondage et peut-être bien, des poids d'estimation (connu sous le nom de poids- g) basés sur l'information auxiliaire. On peut se servir de la méthode de linéarization de Taylor de Binder (1983), afin d'obtenir les écart-types d'estimations de paramètre impliquant les poids d'estimation, tels que ceux qui résultent à l'estimateur de la post - stratification ou de l'ajustement par la régression. Ces écart-types incorporent les poids d'estimation ainsi que les résidus synthétiques obtenus en régressant les composants des fonctions estimantes sur les variables auxiliaires. Pour l'échantillonnage stratifié aléatoire simple, nous obtenons aussi les écart-types en linéarisant l'estimateur de variance jackknife. Nous démontrons comment on pourrait implanter cette méthodologie dans un logiciel qui nous donnerait une approche flexible, unifiée pour calculer des estimations et les écart-types associés.

MOTS CLÉS: Equations estimantes; estimation de variance; linéarization jackknife; linéarization de Taylor

1. INTRODUCTION

Parameters of interest that are estimated in survey sampling are simple or complex. Simple parameters are mostly used for descriptive purposes: totals, ratios, and proportions are included in this class. Complex parameters are used to obtain a better understanding of the relationships that hold within the population of interest: examples of such parameters include regression vectors, logistic regression vectors, and log linear model parameters. The estimators of simple parameters are straightforward to obtain. However, the estimation of complex parameters requires a suitable modification to the generalized linear model approach described by Nelder and Wedderburn (1972). This modification transforms their theory developed in the context of infinite populations to finite populations. Linear or nonlinear parameters of interest, such as population

[1] M.A. Hidiroglou, Business Survey Methods Division, Statistics Canada, Tunney's Pasture, Ottawa, Ontario, K1A 0T6, Canada; J.N.K. Rao, School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, K1S 5B6, Canada; W. Yung, Business Survey Methods Division, Statistics Canada, Tunney's Pasture, Ottawa, Ontario, K1A 0T6, Canada.

means, ratios, linear and logistic regression coefficients can be expressed as solutions to suitable census "estimating equations" (Binder, 1983 Godambe and Thompson, 1986). Parameter estimates are obtained by solving sample estimating equations which involve the design weights as well as possibly calibration weights based on auxiliary information. Using Binder's (1983) Taylor linearization method, we obtain standard errors of parameter estimates involving calibration weights such as those resulting from post-stratification or regression adjustments. The resulting standard errors incorporate the calibration weights as well as synthetic residuals obtained by regressing the components of the estimating functions on the auxiliary variables. For stratified random sampling, we also obtain standard errors by linearizing the jackknife variance estimator.

Section 2 provides the population estimating equations for obtaining the parameter of interest, and gives some examples of how this approach generates commonly known parameters in survey sampling. Section 3 presents the procedure for estimating the parameters of interest via the estimating equation approach, while section 4 offers two alternatives for estimating the variance of the resulting estimators. In Section 5, we display how these results can be applied to a number of post-stratified estimators, and finally a computerization of the proposed procedure is given in Section 6.

2. CENSUS PARAMETERS

We suppose that the finite population U is of size N, and that for each unit k we have data $(y_k, \boldsymbol{x}_k^T)$ where the $\boldsymbol{x}_k$'s are P-dimensional explanatory variables, and the y_k's are response variables. Assume for a given $\boldsymbol{x}$, that the y-value is generated by a random process with mean $E(y_k)=\mu_k=\mu(\boldsymbol{x}_k,\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a P-vector of parameters.

Denote the "working" variance of y_k by $V(y_k)=V_{0k}=\sigma^2 V_0(\mu_k)$ for $k \in U$. A census parameter $\boldsymbol{\theta}_N$ is defined by the solution to the *population estimating equation*

$$S(\boldsymbol{\theta})=\sum_U \boldsymbol{u}_k(\boldsymbol{\theta})=\boldsymbol{0}, \qquad (2.1)$$

where the p-th element of $\boldsymbol{u}_k(\boldsymbol{\theta})$ is

$$u_{kp}(\boldsymbol{\theta})=(\partial\mu_k/\partial\theta_p)\{(y_k-\mu_k)/V_{0k}\} \quad (p=1,\ldots,P)(2.2)$$

The estimating equation approach can be used to generate most of the commonly used census parameters $\boldsymbol{\theta}_N$, e.g., the mean, ratio of two totals, (y and x), linear regression and logistic regression parameters.

2.1 Mean

The model for generating a variable y is given by $E(y_k)=\mu_k=\theta$, $V(y_k)=\sigma^2$, and $Cov(y_k,y_\ell)=0$ for $k\neq\ell=1,\ldots,N$. Using (2.1), this model leads to $u_k(\theta)=y_k-\theta$, $S(\theta)=\Sigma_U\,(y_k-\theta)$ and $\theta_N=N^{-1}\Sigma_U\,y_k=\bar{Y}$.

2.2 Ratio of two totals

The model for generating a ratio of two means is given by y_k is $E(y_k)=\mu=\theta\, x_k$, $V(y_k)=\sigma^2 x_k$, and $Cov(y_k,y_\ell)=0$ for $k\neq\ell$. This leads to $u_k(\theta)=y_k-\theta\, x_k$, $S(\theta)=\Sigma_U\,(y_k-\theta x_k)$, and $\theta_N=\Sigma_U\, y_k/\Sigma_U\, x_k$.

2.3 Linear Regression

The model for generating the parameters of a linear regression of y on x is given by $E(y_k)=\boldsymbol{x}_k^T\boldsymbol{\theta}$, $V(y_k)=\sigma^2$ and $Cov(y_k,y_\ell)=0$ $k\neq\ell$. In this case $\boldsymbol{u}_k(\boldsymbol{\theta})=\boldsymbol{x}_k(y_k-\boldsymbol{x}_k^T\boldsymbol{\theta})$, $S(\boldsymbol{\theta})=\Sigma_k\boldsymbol{x}_k(y_k-\boldsymbol{x}_k^T\boldsymbol{\theta})$, and the census parameter vector given by $\boldsymbol{\theta}_N=\left(\sum_U \boldsymbol{x}_k\boldsymbol{x}_k^T\right)^{-1}\sum_U \boldsymbol{x}_k y_k$.

2.4 Logistic Regression

The model for generating the parameters of logistic regression is given by $E(y_k)=\mu$, where y_k is a dichotomous random variable taking 0 and 1 values, and

$$\mu_k=e^{\theta^T x_k}(1+e^{\theta^T x_k})^{-1}.$$

As a working model for the variance, we take the standard binomial form with $V_{0k}=\mu_k(1-\mu_k)$, so that $\boldsymbol{u}_k=\boldsymbol{x}_k(y_k-\mu_k)$ and $S(\boldsymbol{\theta})=\Sigma_U\boldsymbol{x}_k(y_k-\boldsymbol{x}_k^T\boldsymbol{\theta})$. The parameter $\boldsymbol{\theta}_N$ is implicitly defined by $S(\boldsymbol{\theta})=\boldsymbol{0}$.

3. PARAMETER ESTIMATION

Auxiliary data is commonly used at the estimation stage to improve the precision of sample estimates or to benchmark to known population totals. These methods, known as calibration methods, obtain final or calibration weights, $\tilde{w}_i$, by minimizing a distance measure from the basic design weights, w_i, subject to the restriction that $\Sigma_s\,\tilde{w}_i z_i=\Sigma_U\, z_i$ where z_i is some auxiliary data vector known for all units or for all sampled units and the population total $Z=\Sigma_U z_i$ is known. The resulting calibration weights can be expressed as the product of the basic design weight, w_i, and adjustment factor, a_i (also called g-weight) obtained from the calibration procedure. Note that if the population totals Z are estimated using the calibration weights, one will benchmark to the known population totals Z. Several procedures for achieving this have been given, see Huang and Fuller (1978) and Deville and Särndal (1992).

Generalized Regression (GREG) is the most commonly used calibration procedure. For the GREG procedure, we assume that either auxiliary variables z_i are available for all units in the population U, or only that the population total $Z = \Sigma_U z_i$ is known. The GREG estimator of the total $Y = \Sigma_U y_i$ is given by $\hat{Y} = \Sigma_s \tilde{w}_i y_i$ and the GREG adjustment factor is given by

$$a_i = 1 + z_i^T \left(\Sigma_s w_i z_i z_i^T\right)^{-1} (Z - \hat{Z}) / q_i \qquad (3.1)$$

where $\hat{Z} = \Sigma_s w_i z_i$ and Σ_s denotes the summation over all sampled units $i \in s$ and q_i are specified constants. The estimator, θ, of the population estimating function $S(\theta) = \Sigma_U u_i(\theta)$ is given by $\hat{S}(\theta) = \sum_s \tilde{w}_i u_i(\theta)$, where the calibration weights are given by $\tilde{w}_i = w_i a_i$. The estimator, $\hat{\theta}$, of the census parameter θ_N is obtained by solving $\hat{S}(\theta) = \sum_s \tilde{w}_i u_i(\theta) = 0$. For simple cases such as totals or means, the solution to the estimating equation has a closed form and calculation of $\hat{\theta}$ is therefore straightforward. For more complex cases, it maybe necessary to solve the sample estimating equations iteratively; for example, the r-th step of the Newton-Raphson algorithm is given by

$$\hat{\theta}_r = \hat{\theta}_{r-1} + J(\hat{\theta}_{r-1})^{-1} \hat{S}(\hat{\theta}_{r-1}),$$

where $\hat{\theta}_{r-1}$ is the value of $\hat{\theta}$ obtained at the $(r-1)$-th iteration, $J(\hat{\theta}_{r-1}) = -\partial \hat{S}(\theta) / \partial \theta$ evaluated at $\hat{\theta}_{r-1}$ and $\hat{S}(\hat{\theta}_{r-1})$ is $\hat{S}(\theta)$ evaluated at $\hat{\theta}_{r-1}$. Iterating the Newton-Raphson algorithm to convergence produces the estimate $\hat{\theta}$.

4. VARIANCE ESTIMATION

We next provide two procedures for estimating variances of the estimated parameters that result from the sample estimating equation. One is based on the Taylor linearization approach and the other is jackknife based. Note that the Taylor approach is applicable to general designs, while the jackknife is restricted to particular sample designs. For the purpose of continuity, we will present both approaches using a stratified simple random sample design. Both the Taylor and jackknife approaches are easily amenable to programming. They are part of the building blocks necessary for computing variance estimates of complex parameter estimates in a system such as the Generalized Estimation System (GES) recently developed at Statistics Canada.

4.1 Linearization Variance Estimates

We assume that:

i) $\hat{S}(\theta)$ is distributed normally with mean $S(\theta)$ and variance $V(\hat{S}(\theta))$;
ii) A consistent estimator of $V(\hat{S}(\theta))$ is given by $\hat{V}(\hat{S}(\theta))$.

A consistent estimator of the variance of $\hat{\theta}$ is given by

$$\hat{V}(\hat{\theta}) = \left[J(\hat{\theta})\right]^{-1} \hat{V}(\hat{S}(\hat{\theta})) \left[J(\hat{\theta})\right]^{-1} \qquad (4.1)$$

where $J(\hat{\theta})$ is

$$J(\theta) = \frac{-\partial \hat{S}(\theta)}{\partial \theta^T} = \sum_s \tilde{w}_i \frac{-\partial u_i(\theta)}{\partial \theta^T}$$

evaluated at $\theta = \hat{\theta}$ and $\hat{V}(\hat{S}(\hat{\theta}))$ is $\hat{V}(\hat{S}(\theta))$ evaluated at $\theta = \hat{\theta}$. Since $\hat{S}(\theta)$ is the estimator of a total for $S(\theta)$, the variance estimator $\hat{V}(\hat{S}(\hat{\theta}))$ is obtained from the GREG results for a total, as shown below.

Denoting the estimator of variance of the basic estimator $\hat{Y} = \Sigma_s w_i y_i$ as $v(y_i)$ in operator notation we have

$$\hat{V}(\hat{S}(\hat{\theta})) = v(a_i e_i^*) \qquad (4.2)$$

where $e_i^* = (e_{i1}^*, ..., e_{iP}^*)^T$ with $e_{ip}^* = e_{ip}^*(\hat{\theta})$ and $e_{ip}^*(\hat{\theta})$ is the residual obtained by regressing each component $u_{ip}(\hat{\theta})$ of $u_i(\hat{\theta})$ on the auxiliary variables, z_i. That is,

$$e_{ip}^*(\hat{\theta}) = u_{ip}(\hat{\theta}) - \hat{B}_p(\hat{\theta})^T z_i, \; p = 1, ..., P \qquad (4.3)$$

where

$$\hat{B}_p(\hat{\theta}) = (\Sigma_s w_i z_i z_i^T / q_i)^{-1} (\Sigma_s w_i z_i u_{ip}(\hat{\theta}) / q_i).$$

It is clear from (4.3) that poststratification adjustment may not lead to gain in efficiency if the model residuals $u_{ip}(\hat{\theta})$ are unrelated to z_i. The use of $a_i e_i^*$ instead of e_i^* can be justified along the lines of Särndal, Swensson, and Wretman (1989). Substituting (4.2) into (4.1) we get $\hat{V}(\hat{S}(\hat{\theta}))$ as

$$\hat{V}(\hat{\theta}) = v\left(\left[J(\hat{\theta})\right]^{-1} a_i e_i^*\right) \qquad (4.4)$$

in operator notation.

4.2 Stratified Random Sampling

We assume stratified random sampling with n_h units sampled from N_h units in h-th stratum, h=1, ..., L. We denote the basic weights as $w_h = N_h / n_h$, $i = 1, ..., n_h$; h=1, ..., L and the calibration weights as $\tilde{w}_{hi} = w_{hi} a_{hi}$, where a_{hi} is given by (3.1) with the subscript i changed to hi. The estimated variance of $\hat{Y}=\sum_s w_{hi}\, y_{hi}$ is computed as

$$v(\hat{Y})=\sum_h (1-f_h)\frac{n_h}{n_h-1}\sum_i (w_{hi}y_{hi}-\bar{y}_{wh})(w_{hi}y_{hi}-\bar{y}_{wh})^T$$
$$=v(y_{hi}), \qquad (4.5)$$

where $f_h=n_h/N_h$ and $\bar{y}_{wh}=(1/n_h)\sum_i w_{hi}\, y_{hi}$. It now follows from (4.4) and (4.5) that

$$\hat{V}(\hat{\theta})=v\Big(\big[J(\hat{\theta})\big]^{-1} a_{hi} e^*_{hi}\Big) \qquad (4.6)$$

with the residuals $e^*_{hip}=e^*_{hip}(\hat{\theta})$ and $e^*_{hip}(\hat{\theta})$ is given by (4.3) by changing the subscript i to hi. That is ,

$$e^*_{hip}=u_{hip}(\hat{\theta})-\hat{B}^T_p z_{hi},$$
$$h=1,\cdots,L;\; i=1,\cdots,n_h;\; p=1,\cdots,P,$$

where

$$\hat{B}_p=\Big(\sum_s w_{hi}\, z_{hi} z^T_{hi}/q_{hi}\Big)^{-1}\sum_s w_{hi}\, z_{hi}\, u_{hip}(\hat{\theta})/q_{hi}$$

The operator notation $v([J(\hat{\theta})]^{-1} a_{hi} e^*_{hi})$ indicates that the variance estimator depends only on the error terms, e^*_{hi}, and $J(\hat{\theta})$.

4.3 Jackknife Procedure

To obtain a variance estimator of $\hat{\theta}$, one can also employ resampling techniques such as the jackknife or bootstrap variance estimators. We present the jackknife variance estimator and derive a linearization type variance estimator by approximating the jackknife variance estimator, assuming stratified random sampling.

To calculate the jackknife variance estimator for $\hat{\theta}$, we first define the jackknife weights when the *j-th* unit in the *g-th* stratum is deleted,

$$w_{hi(gj)}=\begin{cases} 0 & (gj)=(hi) \\ \dfrac{n_g}{n_g-1}w_{gi} & h=g, i\neq j \\ w_{hi} & h\neq g \end{cases}$$

The jackknife weight can be interpreted as follows; for the *(gj)-th* unit, the jackknife weight is equal to zero, while for other units in the *g-th* stratum but not in the *j-th* unit, the design weights are adjusted by $n_g/(n_g-1)$. Finally, the weights are unchanged for those units not in the *g-th* stratum. The sample estimating equation when the *(gj)-th* unit is deleted is then

$$\hat{S}_{(gj)}(\theta)=\sum_s \tilde{w}_{hi(gj)} u_{hi}(\theta)=0 \qquad (4.7)$$

where $\tilde{w}_{hi(gj)}$ are the jackknife adjusted final weights. These weights are obtained in the same way as the original final weights, $\tilde{w}_{hi}$, except that the jackknife weights are used instead of the original design weights in the calculation of the adjustment factors a_{hi}.

Now to obtain $\hat{\theta}_{(gj)}$, one can use the Newton Raphson algorithm with equation (4.7) and iterate until convergence, or one can use the one-step jackknife (Lipsitz, Dear and Zhao, 1994). The one-step jackknife simply uses the full sample estimate, $\hat{\theta}$, as the starting point and performs only one step of the Newton-Raphson algorithm with the final weights replaced by the jackknife final weights, $\tilde{w}_{hi(gj)}$. That is,

$$\hat{\theta}_{(gj)}=\hat{\theta}+J_{(gj)}(\hat{\theta})^{-1}\hat{S}_{(gj)}(\hat{\theta})$$

where $J_{(gj)}(\hat{\theta})$ and $\hat{S}_{(gj)}(\hat{\theta})$ are obtained from $J(\hat{\theta})$ and $\hat{S}(\hat{\theta})$ with the final weights replaced by the jackknife final weights. The jackknife variance estimator, using $\hat{\theta}_{(gj)}$, and $(\hat{\theta})$, is given by

$$\hat{V}_J(\hat{\theta})=\sum_g (1-f_g)\left(\frac{n_g-1}{n_g}\right)\sum_j (\hat{\theta}_{(gj)}-\hat{\theta})(\hat{\theta}_{(gj)}-\hat{\theta})^T. \qquad (4.8)$$

To obtain a linearization variance estimator, note that if we have $J_{(gj)}(\hat{\theta})^{-1}\approx J(\hat{\theta})^{-1}$, then the one-step jackknife gives

$$\hat{\theta}_{(gj)}-\hat{\theta}\approx -J(\hat{\theta})^{-1}\hat{S}_{(gj)}(\hat{\theta}) \qquad (4.9)$$

and since $\hat{S}(\hat{\theta})=0$, we have from (4.8) and (4.9)

$$\hat{V}_J(\hat{\theta})\approx\sum_g (1-f_g)\left(\frac{n_g-1}{n_g}\right)\sum_j [J(\hat{\theta})^{-1}\hat{S}_{(gj)}(\hat{\theta})][J(\hat{\theta})^{-1}\hat{S}_{(gj)}(\hat{\theta})]^T$$
$$=J(\hat{\theta})^{-1}\sum_g (1-f_g)\left(\frac{n_g-1}{n_g}\right)\sum_j [\hat{S}_{(gj)}(\hat{\theta})-\hat{S}(\hat{\theta})][\hat{S}_{(gj)}(\hat{\theta})-\hat{S}(\hat{\theta})]^T J(\hat{\theta})^{-1}$$
$$=J(\hat{\theta})^{-1}\hat{V}_J\big[\hat{S}(\hat{\theta})\big]J(\hat{\theta})^{-1} \qquad (4.10)$$

Again, noting that $\hat{S}(\hat{\theta})$ is simply a GREG estimator of the population estimating function, we can approximate $\hat{V}_J\big[\hat{S}(\hat{\theta})\big]$ by the linearized jackknife variance estimator given by

$$\hat{V}_{JL}(\hat{S}(\hat{\theta}))=v(a_{hi}e^*_{hi}), \qquad (4.11)$$

where the *p-th* component of e^*_{hi} is $e_{hip}=u_{hip}(\hat{\theta})-\hat{B}^T_p z_{hi}$, and $\hat{B}_p=(\sum_s w_{hi}z_{hi}z^T_{hi})^{-1}\sum_s w_{hi}z_{hi}u_{hip}(\hat{\theta})$ This result can be obtained by applying techniques in Yung and Rao (1996) who considered the GREG estimator of a total in a stratified multistage framework. Note that the linearized jackknife variance estimator uses g-weighted residuals $a_{hi}e^*_{hi}$. It now follows from (4.10) and (4.11) that, $\hat{V}_J(\hat{\theta})$ can be approximated by the linearized jackknife given by

$$\begin{aligned}\hat{V}_{JL}(\hat{\theta}) &= [J(\hat{\theta})]^{-1}\, v(a_{hi}e^*_{hi})\, [J(\hat{\theta})]^{-1} \\ &= v([J(\hat{\theta})]^{-1}e^*_{hi}).\end{aligned} \qquad (4.12)$$

The estimator (4.12) is identical to the estimator (4.6) obtained using the Taylor method. In Section 5, we consider several specific estimators.

5. SOME APPLICATIONS

The preceding results provide us with a relatively easy procedure for computing variances of estimators of complex parameters, that incorporate auxiliary data in their weights. In this section, we confine to one of the more commonly used weight adjustment procedure, post-stratification. The post-stratified weight adjustment is a particular type of the GREG adjustment.

Let the population U be divided or post-stratified into Q post-strata $U_q, q = 1,\cdots,Q$. Denote the auxiliary data as $z_{hi} = (\delta_{hi1},\cdots,\delta_{hiQ})^T$ where $\delta_{hiq} = 1$ if $(hi) \in U_q$ and 0 otherwise. The adjustment factor for this calibration method is

$$a_{hi} = \sum_q \delta_{hiq} N_q / \hat{N}_q$$

where N_q is the population count for the q-th post-strata and $\hat{N}_q = \Sigma_s w_{hi}\delta_{hiq}$ is an estimate of N_q. The post-stratified estimator of the total is

$$\hat{Y}_{PS} = \sum_s \tilde{w}_{hi} y_{hi} = \sum_q \frac{N_q}{\hat{N}_q} \hat{Y}_q, \text{ where } \hat{Y}_q = \Sigma_s w_{hi} y_{hi} \delta_{hiq}.$$

We now illustrate the Taylor and linearized jackknife methods for obtaining variance estimators with several specific estimating equations. Note that the resulting variance estimates are identical.

5.1 Post-stratified Mean

The sample estimating equation for the post-stratified estimator of the mean is

$$\hat{S}(\theta) = \sum_s \tilde{w}_{hi}(y_{hi} - \theta),$$

with solution

$$\begin{aligned}\hat{\theta} &= \frac{\sum_s \tilde{w}_{hi} y_{hi}}{\sum_s \tilde{w}_{hi}} \\ &= (1/N)\sum_s \tilde{w}_{hi} y_{hi},\end{aligned}$$

since $\Sigma_s \tilde{w}_{hi} = N$ by the benchmarking property of the calibration weights.

To obtain a linearization variance estimator, first note that

$$\begin{aligned}\hat{\boldsymbol{B}} &= \left(\sum_s w_{hi} z_{hi} z_{hi}^T\right)^{-1} \sum_s w_{hi} z_{hi} u_{hi}(\hat{\theta}) \\ &= diag\,(\hat{N}_1^{-1},\dots,\hat{N}_Q^{-1})\left[(\hat{Y}_1,\dots,\hat{Y}_Q)^T - \hat{\theta}(\hat{N}_1,\dots,\hat{N}_Q)^T\right] \\ &= (\bar{y}_1,\dots,\bar{y}_Q)^T - \hat{\theta}(1,\dots,1)^T,\end{aligned}$$

where $\bar{y}_q = \hat{Y}_q/\hat{N}_q$ is an estimate of the q-th post-stratum mean. The error terms are given by

$$\begin{aligned}e^*_{hi} &= u_{hi}(\hat{\theta}) - \hat{\boldsymbol{B}}^T z_{hi} \\ &= y_{hi} - \hat{\theta} - \sum_p \bar{y}_p \delta_{hiq} + \hat{\theta}, \\ &= y_{hi} - \sum_q \bar{y}_q \delta_{hiq}.\end{aligned}$$

Finally note that $J(\hat{\theta}) = N$. The linearization variance estimator is then given by $v(\hat{\theta}) = v(N^{-1}a_{hi}e^*_{hi})$ which corresponds to the well-known variance estimator for the post-stratified estimator.

5.2 The Ratio of Post-stratified Totals

The ratio of two post-stratified estimators is obtained by solving the sample estimating equation

$$\hat{S}(\theta) = \Sigma_s \tilde{w}_{hi}(y_{hi} - \theta x_{hi}) = 0,$$

where the underlying model is $E(y_{hi}) = \theta x_{hi}$, $V(y_{hi}) = \sigma^2 x_{hi}$. The solution to this estimating equation is given by

$$\hat{\theta} = \frac{\sum_q \hat{Y}_q (N_q/\hat{N}_q)}{\sum_q \hat{X}_q (N_q/\hat{N}_q)} = \frac{\hat{Y}_{PS}}{\hat{X}_{PS}}.$$

Also $J(\hat{\theta}) = \Sigma_s \tilde{w}_{hi} x_{hi} = \hat{X}_{PS}$. Next, note that

$$\hat{\boldsymbol{B}} = (\bar{y}_1,\dots,\bar{y}_Q)^T - \hat{\theta}(\bar{x}_1,\dots,\bar{x}_Q)^T,$$

where $\bar{x}_q = \hat{X}_q/\hat{N}_q$. The error terms are then given by

$$\begin{aligned}e^*_{hi} &= y_{hi} - \hat{\theta}x_{hi} - \boldsymbol{B}^T z_{hi} \\ &= y_{hi} - \hat{\theta}x_{hi} - \sum_q \bar{y}_q\delta_{hiq} - \hat{\theta}\sum_q \bar{x}_q\delta_{hiq} \\ &= y_{hi} - \sum_q \bar{y}_q\delta_{hiq} - \hat{\theta}\,(x_{hi} - \sum_q \bar{x}_q\delta_{hiq})\end{aligned}$$

and a linearization variance estimator is given by $v(\hat{\theta}) = v(\hat{X}_{PS}^{-1}\, a_{hi}\, e^*_{hi})$, where the error terms are as defined above.

5.3 Post-Stratified Linear Regression

The parameters of a linear regression of y_{hi} on $\boldsymbol{x}_{hi}$ is obtained by solving the estimating equation

$$\hat{S}(\boldsymbol{\theta}) = \sum_s \tilde{w}_{hi}\boldsymbol{x}_{hi}(y_{hi} - \boldsymbol{x}_{hi}^T\boldsymbol{\theta}) = \sum_s \tilde{w}_{hi}\, u_{hi}(\boldsymbol{\theta}) = \boldsymbol{0}$$

where the underlying model is $E(y_{hi}) = \boldsymbol{x}_{hi}^T\boldsymbol{\theta}$ and $V(y_{hi}) = \sigma^2$. The solution is given by

$$\hat{\boldsymbol{\theta}} = \left(\sum_s w_{hi}\boldsymbol{x}_{hi}\boldsymbol{x}_{hi}^T\right)^{-1}\sum_s w_{hi}\boldsymbol{x}_{hi}\, y_{hi}.$$

To obtain a linearization variance estimator, first we note that $J(\boldsymbol{\theta}) = \Sigma_s \tilde{w}_{hi}\boldsymbol{x}_{hi}\boldsymbol{x}_{hi}^T$. The *p-th* component of the error terms obtained from regressing $\boldsymbol{u}_{hi}(\hat{\boldsymbol{\theta}})$ on the auxiliary data z_{hi} yields $e_{hip}^* = u_{hip}(\hat{\boldsymbol{\theta}}) - \sum_q \bar{u}_{qp}(\hat{\boldsymbol{\theta}})\,\delta_{hiq}$, where $\bar{u}_{qp} = (1/\hat{N}_q)\,\Sigma_s w_{hi} u_{hip}(\hat{\boldsymbol{\theta}})\,\delta_{hiq}$. The linearization variance estimator is then given by $v(\hat{\boldsymbol{\theta}}) = v\left([J(\hat{\boldsymbol{\theta}})]^{-1} a_{hi}\, e_{hi}^*\right)$, where the error terms and $J(\hat{\boldsymbol{\theta}})$ are defined above.

5.4 Post-Stratified Logistic Regression

The sample estimating equation is $\hat{S}(\boldsymbol{\theta}) = \Sigma_s\ \tilde{w}_{hi}\boldsymbol{x}_{hi}(y_{hi} - \mu_{hi})$ where $\mu_{hi} = \exp(\boldsymbol{x}_{hi}^T\boldsymbol{\theta})/\left(1 + \exp(\boldsymbol{x}_{hi}^T\boldsymbol{\theta})\right) = \mu_{hi}(\hat{\boldsymbol{\theta}})$. The solution to this estimating equation can be obtained using the Newton-Raphson algorithm. Note that $J(\hat{\boldsymbol{\theta}}) = \Sigma_s \tilde{w}_{hi}\hat{\mu}_{hi}(1 - \hat{\mu}_{hi})\boldsymbol{x}_{hi}\boldsymbol{x}_{hi}^T$ where $\hat{\mu}_{hi} = \mu_{hi}(\hat{\boldsymbol{\theta}})$. The *p-th* component of the error term is given by

$$e_{hip}^* = u_{hip}(\hat{\boldsymbol{\theta}}) - \Sigma_q \bar{u}_{qp}(\hat{\boldsymbol{\theta}})\,\delta_{hiq},$$

where $\bar{u}_{qp}(\hat{\boldsymbol{\theta}}) = (1/\hat{N}_q)\,\Sigma_s w_{hi} u_{hip}(\hat{\boldsymbol{\theta}})\,\delta_{hiq}$ A linearization variance estimator is then given by

$$\hat{V}(\hat{\boldsymbol{\theta}}) = v\left([J(\hat{\boldsymbol{\theta}})]^{-1}\, a_{hi}\, e_{hi}^*\right),$$

where the error terms and $J(\hat{\boldsymbol{\theta}})$ are given above.

6. COMPUTER IMPLEMENTATION

Some of the software that specializes in variance estimation for complex surveys have been implemented with the use of auxiliary data. They include the Generalized Estimation System (GES) from Statistics Canada, CLAN from Statistics Sweden, POULPE from INSEE France and BASCULA from the Netherlands. SUDAAN is currently the only software that incorporates the use of estimating equations although it does not use auxiliary information.

GES can compute the estimated variances of a ratio of two estimated totals , where each one may have been calibrated. The post-stratified estimator is the simplest operation of this type. GES is currently set up in modular form with respect to design, but not with respect to smooth functions of totals that incorporate auxiliary data. The design-modular form consists of variance operators that can handle stratified unistage cluster or element sample designs. Multi-stage sample designs, with auxiliary data incorporated at each stage, can also be handled with GES provided that the sampling is stratified simple random sampling at each stage.

Utilizing the framework of estimating equations as presented in this paper, GES can be modified to consist of the following four main modules under GREG calibration:

1. A module to perform the weighting and to calculate the adjustment factors, a_i given by (3.1).
2. A module to estimate $\hat{\boldsymbol{\theta}}$. Use of the Newton-Raphson algorithm would automatically calculate $J(\hat{\boldsymbol{\theta}})$.
3. A module to estimate the residuals, e_i^* given by (4.3).
4. A module to estimate the variance of a total according to the sample design used. The use of the operator notation would allow us to extend the estimator of the variance of a total to an estimator of the variance of $\hat{\boldsymbol{\theta}}$ as
$$\hat{V}(\hat{\boldsymbol{\theta}}) = v([J(\hat{\boldsymbol{\theta}})]^{-1} a_i e_i^*).$$

The above steps can easily be extended to include model groups and domain estimation (Estevao, Hidiroglou, and Särndal, 1995).

7. CONCLUSIONS

Estimates of parameters of interest such as means, ratios of totals and linear or logistic regressions parameters can be obtained by solving suitable sample estimating equations. This paper has provided linearization type variance estimators for a general estimating equation using both the Taylor procedure and a jackknife linearization method. Results are spelled out for a stratified simple random sample design. Extension to the commonly used stratified multi-stage design will be reported in a separate paper. The computer implementation of the proposed method has been outlined with the goal of including this new methodology in Statistics Canada's Generalized Estimation System (GES).

REFERENCES

Andersson, C., and Nordberg, L. (1994). A Method for Variance Estimation of Non-Linear Functions of Totals in Surveys. Theory and Software Implementation. *Journal of Official Statistics*, **10**, 395 - 405.

Binder, D.A. (1983). On the variance of asymptotically normal estimators from complex survey. *International Statistical Review*, **51**, 279-292.

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

Estevao, V., Hidiroglou, M.A., and Särndal, C.-E. (1995)." Methodological Principles for a Generalized Estimation System at Statistics Canada". *Journal of Official* Statistics, Vol.11, No.2, 181-204.

Godambe, V.P. and Thompson, M.E. (1986). Parameters of superpopulation and survey population: their relationship and estimation. *International Statistical Review*, **54**, 127-138.

Huang, E. T. and Fuller, W. A. (1978). Nonnegative regression estimation for survey data. *Proceedings of the Social Statistics Section*, American Statistical Association 1978, 300 - 303.

Liang , K. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.

Lipsitz, S.R., Dear, K.B.G., and Zhao, L. (1994). Jackknife estimators of variance for parameter estimates from estimating equations with applications to clustered survival data. *Biometrics*, **50**, 842-846.

Nelder, J.A. and Wedderburn, R.W.M. (1992). Generalized linear models. *Journal of the Royal Statistical Society A*, **135**, 370-384.

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.

Yung, W. and Rao, J.N.K. (1996). Jackknife linearization variance estimators under stratified multi-stage sampling. *Survey Methodology,* **22**, 23-31.