VARIANCE ESTIMATION AND QUASI-SCORE TESTS FOR COMPLEX SURVEYS USING ESTIMATING EQUATIONS

M.A. HIDIROGLOU, J.N.K. RAO AND W. YUNG
Mike Hidiroglou, Statistics Canada, 11-A R.H. Coats Building, Tunney's Pasture, Ottawa, Canada, K1A-0T6

Key Words: Estimating equations, jackknife linearization, Taylor linearization, variance estimation.

1. Introduction

Parameters of interest that are estimated in survey sampling are simple or complex. Simple parameters such as totals, ratios, and proportions are used mostly for descriptive purposes. 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 parameters of log linear models. 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 means, ratios and linear or 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 estimation weights based on auxiliary information. Using Binder's (1983) Taylor linearization method, we obtain standard errors of parameter estimates involving design and estimation weights, such as those resulting from post-stratification or regression adjustments. The resulting standard errors incorporate the estimation weights as well as synthetic residuals obtained by regressing the components of the estimating functions on the auxiliary variables. For stratified random sampling and stratified multistage sampling we also obtain standard errors by linearizing a jackknife variance estimator. Finally, analogues to C.R. Rao's score tests that take account of survey design features are developed.

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. Section 6 develops analogues of C.R. Rao's score tests that take account of survey design features. Finally, a computer implementation of the proposed method is given in Section 7.

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 equations*

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

where Σ_U denotes the summation over the finite population U and the p-th element of $\boldsymbol{u}_k(\boldsymbol{\theta})$ is

$$u_{kp}(\boldsymbol{\theta}) = \left(\frac{\partial \mu_k}{\partial \theta_p}\right)\left\{\frac{(y_k - \mu_k)}{V_{0k}}\right\}, \tag{2.2}$$

where $p=(1,...,P)$.

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

The model for generating the mean of 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 $k \neq \ell = 1,...,N$. Using (2.1) and (2.2), this model leads to $u_k(\theta) = y_k - \theta$, $S(\theta) = \Sigma_U (y_k - \theta)$ and $\theta_N = N^{-1} \Sigma_U y_k = \bar{Y}$, the mean of y. The model for generating a ratio of two totals is given by $E(y_k) = \mu = \theta x_k$, $V(y_k) = \sigma^2 x_k$, 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 $\theta_N = \Sigma_U y_k / \Sigma_U x_k$, the ratio of totals of y and x. 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 $\boldsymbol{u}_k(\boldsymbol{\theta}) = \boldsymbol{x}_k(y_k - \boldsymbol{x}_k^T \boldsymbol{\theta})$, $\boldsymbol{S}(\boldsymbol{\theta}) = \Sigma_k \boldsymbol{x}_k(y_k - \boldsymbol{x}_k^T \boldsymbol{\theta})$, and the census parameter vector is given by $\boldsymbol{\theta}_N = \left(\sum_U \boldsymbol{x}_k \boldsymbol{x}_k^T\right)^{-1} \sum_U \boldsymbol{x}_k y_k$. The model for generating the parameters of a logistic regression is given by

$E(y_k) = \mu_k$, where y_k is a dichotomous random variable taking 0 and 1 values, and $\mu_k = \exp(\theta^T x_k)\left[1 + \exp(\theta^T x_k)\right]^{-1}$. As a working model for the variance, we take the standard binomial form with $V_{0k} = \mu_k(1 - \mu_k)$, so that $u_k(\theta) = x_k(y_k - \mu_k)$ and $S(\theta) = \Sigma_U x_k(y_k - x_k^T\theta)$. The parameter θ_N is implicitly defined by $S(\theta) = 0$.

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 design weights, w_i, subject to the restriction that $\Sigma_s \tilde{w}_i z_i = \Sigma_U z_i$ where Σ_s denotes the summation over a sample of units, z_i is some auxiliary data vector known for all populations units, or for all sampled units and the population total $Z = \Sigma_U z_i$ is known. The resulting calibration weight can be expressed as the product of the design weight, w_i, and estimation weight, a_i, obtained from the calibration procedure, i.e. $\tilde{w}_i = w_i a_i$. Note that if the population totals are estimated using the calibration weights, one will benchmark to the known population totals, Z.

Generalized Regression (GREG) is the most commonly used calibration procedure. The GREG estimator of the total $Y = \Sigma_U y_i$ is given by $\hat{Y}_G = \Sigma_s \tilde{w}_i y_i$ and the GREG estimation weight is given by

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

where $\hat{Z} = \Sigma_s w_i z_i$ and the q_i's are specified constants. The estimator of the population estimating function $S(\theta) = \Sigma_U u_i(\theta)$ is given by $\hat{S}(\theta) = \Sigma_s \tilde{w}_i u_i(\theta)$. The estimator, $\hat{\theta}$, of the census parameter θ_N is obtained by solving $\hat{S}(\theta) = \Sigma_s \tilde{w}_i u_i(\theta) = 0$. For simple cases such as totals or means, the solution to the estimating equations has a closed form and the calculation of $\hat{\theta}$ is therefore straightforward. For more complex cases, such as logistic regression, it may be necessary to solve the sample estimating equations iteratively; for example, the *r-th* step of the Newton-Raphson algorithm is given by

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

where $\hat{\theta}_{r-1}$ is the value of $\hat{\theta}$ obtained at the $(r-1)$*-th* iteration, $J(\hat{\theta}_{r-1}) = -\partial\hat{S}(\theta)/\partial\theta^T$ 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 equations. One method uses the Taylor linearization approach and the other is based on the jackknife method. Note that the Taylor linearization approach is applicable to general designs, while the jackknife is restricted to particular sample designs. Since both approaches are easily amenable to programming, they could be 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 Taylor Linearization

We assume that (i) $\hat{S}(\theta)$ is asymptotically normal with mean $S(\theta)$ and covariance matrix $V(\hat{S}(\theta))$; (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 then given by

$$\hat{V}(\hat{\theta}) = J^{-1}(\hat{\theta})\,\hat{V}(\hat{S}(\hat{\theta}))\,J^{-1}(\hat{\theta}) \qquad (4.1)$$

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 an estimator of $\hat{V}(\hat{S}(\theta))$ (Binder, 1983). Since $\hat{S}(\hat{\theta})$ is the GREG estimator of the total $S(\theta) = \Sigma_U u_i(\theta)$, the variance estimator $\hat{V}(\hat{S}(\hat{\theta}))$ is readily obtained from the GREG results for a total, as shown below.

Using operator notation, let $v(y_i)$ denote the estimator of the covariance matrix of the basic estimator $\hat{Y} = \Sigma_s w_i y_i$ where $y_i = (y_{i1}, \ldots, y_{iP})^T$. Then an estimator of the covariance matrix of the GREG estimator $\hat{Y}_G = \Sigma_s \tilde{w}_i y_i$ is given by

$$\hat{V}(\hat{Y}_G) = v(a_i e_i), \qquad (4.2)$$

where $e_i = (e_{i1}, \ldots, e_{iP})^T$ with e_{ip} $(p = 1, \ldots P)$ denoting the residual obtained by regressing the *p-th* component y_{ip} on the auxiliary variables z_i. That is, $e_{ip} = y_{ip} - \hat{B}_p^T z_i$ where

$$\hat{B}_p = \left(\Sigma_s w_i z_i z_i^T/q_i\right)^{-1}\left(\Sigma_s w_i z_i y_{ip}/q_i\right),$$

and the q_i's are specified constants motivated by a "working" model. The use of $a_i e_i$ in (4.2) instead of e_i can be justified along the lines of Särndal, Swensson and Wretman (1989).

Returning to the covariance matrix of $\hat{S}(\hat{\theta})$, it readily follows from (4.2) that an estimator of the covariance matrix of $\hat{S}(\hat{\theta})$ is given by

$$\hat{V}(\hat{S}(\hat{\theta})) = v(a_i e_i^*) \qquad (4.3)$$

where $e_i^* = (e_{i1}^*, \ldots, e_{iP}^*)^T$ with $e_{ip}^* = e_{ip}^*(\hat{\theta})$. The residuals, $e_{ip}^*(\hat{\theta})$, are obtained by regressing the *p-th* component,

$u_{ip}(\hat{\theta})$, of $u_i(\hat{\theta})$ on the auxiliary variables, z_i. That is,

$$e^*_{ip}(\hat{\theta}) = u_{ip}(\hat{\theta}) - \hat{B}_p(\hat{\theta})^T z_i, \; p = 1, ..., P \qquad (4.4)$$

where

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

It is clear from (4.4) that calibration estimation may not lead to gains in efficiency if the model residuals $u_{ip}(\hat{\theta})$ are unrelated to z_i. Substituting (4.3) into (4.1) we get $\hat{V}(\hat{\theta})$ as

$$\hat{V}(\hat{\theta}) = v\left(J^{-1}(\hat{\theta}) a_i e^*_i \right) \qquad (4.5)$$

in operator notation.

We now spell out the general formula (4.5) for two commonly used designs: stratified random sampling and stratified multistage sampling.

4.1.2 Stratified Random Sampling

Stratified random sampling is commonly used in establishment surveys based on list frames; for example, in business surveys at Statistics Canada. Suppose n_h units are selected by simple random sampling from N_h units in the *h-th* stratum, $h=1, ..., L$ independently across strata. We denote the basic weight attached to the *i-th* sample unit in the *h-th* stratum as $w_{hi} = N_h / n_h$, $i=1, ..., n_h$; $h=1, ..., L$, 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 covariance matrix of $\hat{Y} = \sum_s w_{hi} y_{hi}$ is given by

$$\hat{V}(\hat{Y}) = \sum_h \frac{(1-f_h) n_h}{n_h - 1} \sum_i \times (w_{hi} y_{hi} - \bar{y}_h)(w_{hi} y_{hi} - \bar{y}_h)^T = v(y_{hi}), \qquad (4.6)$$

where $f_h = n_h / N_h$ and $\bar{y}_h = (1/n_h) \sum_i w_{hi} y_{hi}$. It now follows from (4.5) and (4.6) that

$$\hat{V}(\hat{\theta}) = v\left(J^{-1}(\hat{\theta}) a_{hi} e^*_{hi} \right), \qquad (4.7)$$

where e^*_{hip} is given by (4.4) by changing the subscript i to hi.

4.1.3 Stratified Multistage Sampling

Stratified multistage sampling is commonly used in large socio-economic surveys; for example, in the Canadian Labour Force Survey. We focus on designs with large number of strata, L, and relatively few primary sampling units (clusters), $n_h (\geq 2)$, sampled within each stratum h ($h=1, ..., L$). We assume that subsampling within sampled clusters i ($=1, ..., n_h$) is performed to ensure unbiased estimation of cluster totals. Let m_{hi} be the number of ultimate units sampled from the *(hi)-th* sample cluster. The design weights w_{hik} attached to the sample units (*hik*) are calibrated to satisfy $\sum_s \tilde{w}_{hik} z_{hik} = \sum_U z_{hik}$, the known totals of auxiliary variables z. The resulting calibration weights are $\tilde{w}_{hik} = w_{hik} a_{hik}$, where a_{hik} is given by (3.1) with the subscript i changed to *hik*.

For variance estimation, the clusters are assumed to be sampled with replacement. The estimated covariance matrix of $\hat{Y} = \sum_s w_{hik} y_{hik}$ is then given by

$$\hat{V}(\hat{Y}) = \sum_h \frac{n_h}{(n_h - 1)} \sum_i (y_{hi} - \bar{y}_h)(y_{hi} - \bar{y}_h)^T = v(y_{hi},) \qquad (4.8)$$

where $y_{hi} = \sum_k (n_h w_{hik}) y_{hik}$ and $\bar{y}_h = (1/n_h) \sum_i y_{hi}$. Note that $\hat{V}(\hat{Y})$ depends only on the cluster totals y_{hi}. The estimator (4.8) generally leads to overestimation, but the relative bias is likely to be small if the first-stage sampling fractions are small.

It readily follows from (4.5) and (4.8) that

$$\hat{V}(\hat{\theta}) = v\left(J^{-1}(\hat{\theta}) \tilde{e}_{hi} \right), \qquad (4.9)$$

where $\tilde{e}_{hi} = (\tilde{e}_{hi1}, ..., \tilde{e}_{hiP})^T$, $\tilde{e}_{hip} = \sum_k (n_h \tilde{w}_{hik}) e^*_{hikp}$, and e^*_{hikp} is given by (4.4) by changing the subscript i to *hik*.

4.2 Jackknife Linearization

As noted earlier, the Taylor linearization approach is applicable to general designs, whereas the jackknife method is restricted to particular designs. We now present the jackknife approach for the two commonly used designs considered in subsections 4.1.2 and 4.1.3.

4.2.1 Stratified Random Sampling

To obtain a variance estimator of $\hat{\theta}$, one can also use resampling techniques such as the jackknife or bootstrap variance estimators. We present a 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* sample unit in the *g-th* stratum is deleted as $w_{hi(gj)} = 0$ if $(hi)=(gj)$; $= n_g/(n_g - 1) w_{gi}$ if $h=g$, $i \neq j$; $= w_{hi}$ if $h \neq g$.

The sample estimating equations when the *(gj)-th* unit is deleted are then given by

$$\hat{S}_{(gj)}(\theta) = \sum_s \tilde{w}_{hi(gj)} u_{hi}(\theta) \qquad (4.10)$$

where $\tilde{w}_{hi(gj)}$ are the jackknife adjusted calibration

weights. These weights are obtained in the same way as the original calibration weights, $\tilde{w}_{hi}$, except that the jackknife weights are used instead of the design weights, w_{hi}, in the calculation of the estimation weights a_{hi}.

Now to obtain the solution, $\hat{\boldsymbol{\theta}}_{(gj)}$, from (4.10), one can use the Newton Raphson algorithm with (4.10) 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, $\hat{\boldsymbol{\theta}}$, as the starting point and performs only one step of the Newton-Raphson algorithm with the calibration weights replaced by the jackknife adjusted calibration weights, $\tilde{w}_{hi(gj)}$. That is,

$$\hat{\boldsymbol{\theta}}_{(gj)} = \hat{\boldsymbol{\theta}} + \boldsymbol{J}_{(gj)}^{-1}(\hat{\boldsymbol{\theta}})\, \hat{\boldsymbol{S}}_{(gj)}(\hat{\boldsymbol{\theta}}) \tag{4.11}$$

where $\boldsymbol{J}_{(gj)}(\hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{S}}_{(gj)}(\hat{\boldsymbol{\theta}})$ are obtained from $\boldsymbol{J}(\hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{S}}(\hat{\boldsymbol{\theta}})$ with $\tilde{w}_{hi}$ replaced by $\tilde{w}_{hi(gj)}$. A jackknife estimator of the covariance matrix of $\hat{\boldsymbol{\theta}}$ is given by

$$\hat{V}_J(\hat{\boldsymbol{\theta}}) = \sum_g (1-f_g)\left(\frac{n_g-1}{n_g}\right) \times \sum_j (\hat{\boldsymbol{\theta}}_{(gj)}-\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{(gj)}-\hat{\boldsymbol{\theta}})^T. \tag{4.12}$$

To obtain a linearization variance estimator, we note that $\boldsymbol{J}_{(gj)}(\hat{\boldsymbol{\theta}})^{-1} \approx \boldsymbol{J}(\hat{\boldsymbol{\theta}})^{-1}$ and $\hat{\boldsymbol{S}}(\hat{\boldsymbol{\theta}})=\mathbf{0}$ and hence from (4.11),

$$\hat{\boldsymbol{\theta}}_{(gj)}-\hat{\boldsymbol{\theta}} \approx \boldsymbol{J}^{-1}(\hat{\boldsymbol{\theta}})\left[\hat{\boldsymbol{S}}_{(gj)}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{S}}(\hat{\boldsymbol{\theta}})\right]. \tag{4.13}$$

It now follows from (4.12) and (4.13) that

$$\begin{aligned}\hat{V}_J(\hat{\boldsymbol{\theta}}) &\approx \boldsymbol{J}^{-1}(\hat{\boldsymbol{\theta}})\Big(\sum_g (1-f_g)\left[(n_g-1)/n_g\right]\sum_j \\ &\quad \times \left[\hat{\boldsymbol{S}}_{(gj)}(\hat{\boldsymbol{\theta}})-\hat{\boldsymbol{S}}(\hat{\boldsymbol{\theta}})\right]\left[\hat{\boldsymbol{S}}_{(gj)}(\hat{\boldsymbol{\theta}})-\hat{\boldsymbol{S}}(\hat{\boldsymbol{\theta}})\right]^T\Big)\boldsymbol{J}^{-1}(\hat{\boldsymbol{\theta}}) \\ &= \boldsymbol{J}^{-1}(\hat{\boldsymbol{\theta}})\, \hat{V}_J\left[\hat{\boldsymbol{S}}(\hat{\boldsymbol{\theta}})\right]\boldsymbol{J}^{-1}(\hat{\boldsymbol{\theta}}),\end{aligned} \tag{4.14}$$

where $\hat{V}_J\left[\hat{\boldsymbol{S}}(\hat{\boldsymbol{\theta}})\right]$ is the jackknife estimator of $\hat{\boldsymbol{S}}(\hat{\boldsymbol{\theta}}) = \sum_s \tilde{w}_{hi}\, \boldsymbol{u}_{hi}(\hat{\boldsymbol{\theta}})$. Again, noting that $\hat{\boldsymbol{S}}(\hat{\boldsymbol{\theta}})$ is simply a GREG estimator, we can approximate $\hat{V}_J\left[\hat{\boldsymbol{S}}(\hat{\boldsymbol{\theta}})\right]$ by the linearized jackknife variance estimator given by

$$\hat{V}_{JL}\left(\hat{\boldsymbol{S}}(\hat{\boldsymbol{\theta}})\right) = v(a_{hi}\boldsymbol{e}_{hi}^*), \tag{4.15}$$

where the *p-th* component of $\boldsymbol{e}_{hi}^*$ is $e_{hip}^* = u_{hip}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{B}}_p(\hat{\boldsymbol{\theta}})^T \boldsymbol{z}_{hi}$ and

$$\hat{\boldsymbol{B}}_p(\hat{\boldsymbol{\theta}}) = \left(\sum_s w_{hi}\boldsymbol{z}_{hi}\boldsymbol{z}_{hi}^T\right)^{-1}\sum_s w_{hi}\boldsymbol{z}_{hi}\boldsymbol{u}_{hip}(\hat{\boldsymbol{\theta}}).$$

This result can be obtained by applying techniques in Yung and Rao (1996) who considered the GREG estimator, $\hat{Y}_G$, in a stratified multistage framework. Note that the linearized jackknife variance estimator uses estimation weighted residuals $a_{hi}\boldsymbol{e}_{hi}^*$. It now follows from (4.14) and (4.15) that $\hat{V}_J(\hat{\boldsymbol{\theta}})$ can be approximated by the linearized jackknife given by

$$\begin{aligned}\hat{V}_{JL}(\hat{\boldsymbol{\theta}}) &= \boldsymbol{J}^{-1}(\hat{\boldsymbol{\theta}})\, v(a_{hi}\boldsymbol{e}_{hi}^*)\ \boldsymbol{J}^{-1}(\hat{\boldsymbol{\theta}}) \\ &= v\left(\boldsymbol{J}^{-1}(\hat{\boldsymbol{\theta}})\, a_{hi}\boldsymbol{e}_{hi}^*\right).\end{aligned} \tag{4.16}$$

The estimator (4.16) is identical to the estimator (4.7) obtained using the Taylor method.

4.2.2 Stratified Multistage Sampling

The jackknife weights $w_{hik(gj)}$ when the *(gj)-th* sample cluster is deleted are given by $w_{hik(gj)} = 0$ if $(hi)=(gj)$; $= n_g/(n_g-1)\, w_{gik}$ if $h=g$, $i\neq j$; $= w_{hik}$ if $h\neq g$. We calculate $\hat{\boldsymbol{\theta}}_{(gj)}$, as in Section 4.2.1, using the one-step Newton-Raphson algorithm. A jackknife estimator of the covariance matrix of $\hat{\boldsymbol{\theta}}$ is given by

$$\hat{V}_J(\hat{\boldsymbol{\theta}}) = \sum_g \frac{n_g-1}{n_g}\sum_j (\hat{\boldsymbol{\theta}}_{(gj)}-\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{(gj)}-\hat{\boldsymbol{\theta}})^T. \tag{4.17}$$

Now, following along the lines of Section 4.2.1, we get a jackknife linearization estimator $\hat{V}_{JL}(\hat{\boldsymbol{\theta}})$ from (4.17). It is identical to the Taylor linearization estimator (4.9).

5. Some Applications

The preceding results provide us with a relatively easy procedure for computing variance estimators of complex parameters, that incorporate auxiliary data in their weights. In this section, we confine ourselves to poststratification, a commonly used calibration method. The post-stratified weight is a particular type of the GREG calibration weight. Let the population U consist of Q post-strata U_q, $q = 1, \cdots, Q$. Denote the auxiliary data as $\boldsymbol{z}_i = (\delta_{i1}, \ldots, \delta_{iQ})^T$ where $\delta_{iq} = 1$ if $(i) \in U_q$ and 0 otherwise. The estimation weight, a_i, reduces to $a_i = \sum_q \delta_{iq} N_q / \hat{N}_q$ where N_q is the known population count for the *q-th* post-strata and $\hat{N}_q = \sum_s w_i \delta_{iq}$ is an estimator of N_q. The post-stratified estimator of the total Y is $\hat{Y}_{PS} = \sum_s \tilde{w}_i y_i = \sum_q (N_q/\hat{N}_q)\, \hat{Y}_q$, where $\hat{Y}_q = \sum_s w_i \delta_{iq} y_i$ is an estimator of the *q-th* poststratum total Y_q.

5.1 Post-stratified Mean

The sample estimating equation for the mean, θ, is $\hat{S}(\theta) = \sum_s \tilde{w}_i (y_i - \theta)$, with the solution $\hat{\theta} = \sum_s \tilde{w}_i y_i / \sum_s \tilde{w}_i = (1/N)\sum_s \tilde{w}_i y_i$ since $\sum_s \tilde{w}_i = N$ by the benchmarking property of the calibration weights. Further, $e_i^* = y_i - \sum_q \delta_{iq}\, \bar{y}_q$ and $J(\hat{\theta}) = N$, where $\bar{y}_q = \hat{Y}_q / \hat{N}_q$ is the estimator of the *q-th* poststratum mean $\bar{Y}_q = Y_q / N_q$. The linearization variance estimator is given by $\hat{V}(\hat{\theta}) = v(N^{-1} a_i e_i^*)$ which agrees with a well-known

variance estimator for the post-stratified estimator (Rao, 1985).

5.2 Ratio of Post-stratified Totals

The ratio of two post-stratified estimators is obtained by solving the sample estimating equation $\hat{S}(\theta) = \sum_s \tilde{w}_i (y_i - \theta x_i) = 0$. The solution to this estimating equation is given by $\hat{\theta} = \hat{Y}_{PS} / \hat{X}_{PS}$ where $\hat{X}_{PS}$ is the post-stratified estimator of the total X. Further, $e_i^* = y_i - \sum_q \delta_{iq} \bar{y}_q - \hat{\theta}(x_i - \sum_q \delta_{iq} \bar{x}_q)$ and $J(\hat{\theta}) = \hat{X}_{PS}$, where $\bar{x}_q = \hat{X}_q / \hat{N}_q$. The linearization variance estimator is given by $\hat{V}(\hat{\theta}) = v(\hat{X}_{PS}^{-1} a_i e_i^*)$.

5.3 Post-Stratified Linear Regression

The sample estimating equations for the linear regression parameters are given by

$$\hat{S}(\theta) = \sum_s \tilde{w}_i \boldsymbol{x}_i (y_i - \boldsymbol{x}_i^T \theta) = \sum_s \tilde{w}_i u_i(\theta) = \mathbf{0},$$

with the solution

$$\hat{\theta} = \left(\sum_s w_i \boldsymbol{x}_i \boldsymbol{x}_i^T \right)^{-1} \sum_s w_i \boldsymbol{x}_i y_i.$$

Further, $e_{ip}^* = u_{ip}(\hat{\theta}) - \sum_q \delta_{iq} \bar{u}_{qp}$ and $\boldsymbol{J}(\hat{\theta}) = \sum_s \tilde{w}_i \boldsymbol{x}_i \boldsymbol{x}_i^T$, where $\bar{u}_{qp} = \left[\sum_s w_i \delta_{ip} u_{ip}(\hat{\theta}) \right] / \hat{N}_q$. The linearization estimator is given by $\hat{V}(\hat{\theta}) = v\left(\boldsymbol{J}^{-1}(\hat{\theta}) a_i e_i^* \right)$.

5.4 Post-Stratified Logistic Regression

The sample estimating equations are given by $\hat{\boldsymbol{S}}(\theta) = \sum_s \tilde{w}_i \boldsymbol{x}_i (y_i - \mu_i) = \sum_s \boldsymbol{u}_i(\theta) = \mathbf{0}$, where $\mu_i(\theta) = \exp(\boldsymbol{x}_i^T \theta) / \left(1 + \exp(\boldsymbol{x}_i^T \theta)\right)$. Further

$$e_{ip}^* = u_{ip}(\hat{\theta}) - \sum_q \delta_{iq} \bar{u}_{qp}$$

and $\boldsymbol{J}(\hat{\theta}) = \sum_s \tilde{w}_i \boldsymbol{x}_i \boldsymbol{x}_i^T [\hat{\mu}_i (1 - \hat{\mu}_i)]$, where $\hat{\mu}_i = \mu_i(\hat{\theta})$. The linearization estimator is given by $\hat{V}(\hat{\theta}) = v\left(\boldsymbol{J}^{-1}(\hat{\theta}) a_i e_i^* \right)$.

6. Quasi-Score Tests

We now turn to the problem of testing null hypotheses of the form H_0: $\theta_2 = \theta_{20}$, where $\theta = \left(\theta_1^T, \theta_2^T \right)^T$ with θ_2 being a $R \times 1$ vector. One approach is to base the test of H_0 on the corresponding Wald Statistic, using $\hat{\theta}$ and $\hat{V}(\hat{\theta})$ given by (4.5), but Wald tests have the following drawbacks: (i) The full model has to be fitted to get $\hat{\theta}$; (ii) Wald tests are not invariant to reparameterizations. C.R. Rao's (1947) score tests are free of both limitations and permit extensions to complex problems. We need only fit the simpler null model, which is a considerable advantage if the full model contains a large number of parameters as will be the case, for example, with a factorial structure of explanatory variables containing a large number of interactions. Bera and Bilias (1998) provide an excellent account of Rao's score test and its extensions in the context of econometric modelling.

Rao, Scott and Skinner (1998) extended Rao's score test to survey data by developing quasi-score tests that take account of survey design features. However, they did not fully investigate the effects of calibration. In this section, we give a brief account of quasi-score tests when calibration weights, $\tilde{w}_i$, are used. Let $\hat{\boldsymbol{S}}(\theta) = \left(\hat{\boldsymbol{S}}_1(\theta)^T, \hat{\boldsymbol{S}}_2(\theta)^T \right)^T$ be the partition of $\hat{\boldsymbol{S}}(\theta)$ corresponding to the partition of θ where $\hat{\boldsymbol{S}}_t(\theta) = \sum_s \tilde{w}_i \boldsymbol{u}_{ti}(\theta)$, $t=1,2$. Also, let $\tilde{\theta} = \left(\tilde{\theta}_1^T, \theta_{20}^T \right)^T$ be the solution of $\hat{\boldsymbol{S}}_1(\theta^*) = \mathbf{0}$, where $\theta^* = \left(\theta_1^T, \theta_{20}^T \right)^T$. The analogue of Rao's score test, which we call the quasi-score test, is based on the statistic

$$QS_R = \tilde{\boldsymbol{S}}_2^T \left[\hat{V}(\tilde{\boldsymbol{S}}_2) \right]^{-1} \tilde{\boldsymbol{S}}_2 \qquad (6.1)$$

where $\tilde{\boldsymbol{S}}_2 = \hat{\boldsymbol{S}}_2(\tilde{\theta})$ and $\hat{V}(\tilde{\boldsymbol{S}}_2)$ is a consistent estimator of $V(\tilde{\boldsymbol{S}}_2)$.

Following Rao, Scott and Skinner (1998), we can write

$$\tilde{\boldsymbol{S}}_2 \approx \sum_s \tilde{w}_i \boldsymbol{u}_{2.1i}(\theta^*),$$

where $\boldsymbol{u}_{2.1i}(\theta^*) = \boldsymbol{u}_{2i}(\theta^*) - \boldsymbol{I}_{21}^* \boldsymbol{I}_{11}^{*-1} \boldsymbol{u}_{1i}(\theta^*)$ and $\boldsymbol{I}^* = \boldsymbol{I}(\theta^*) = E\left[\boldsymbol{J}(\theta^*) \right]$ is partitioned as

$$\boldsymbol{I}^* = \begin{bmatrix} \boldsymbol{I}_{11}^* & \boldsymbol{I}_{12}^* \\ \boldsymbol{I}_{21}^* & \boldsymbol{I}_{22}^* \end{bmatrix}.$$

Since $\tilde{\boldsymbol{S}}_2$ is approximated by an estimated total, it follows that $\tilde{\boldsymbol{S}}_2$ is asymptotically a R-variate normal with mean $\mathbf{0}$ and covariance matrix $V(\tilde{\boldsymbol{S}}_2)$ under H_0. Therefore, the quasi-score statistic QS_R is asymptotically a χ_R^2 variable under H_0. Following (4.3) and (4.4), a Taylor linearization estimator of $V(\tilde{\boldsymbol{S}}_2)$ is given by $\hat{V}(\tilde{\boldsymbol{S}}_2) = v(a_i \boldsymbol{e}_{2.1i}^*)$, where $\boldsymbol{e}_{2.1i}^*$ is a R-vector with elements $\boldsymbol{e}_{2.1iq}^* = u_{2.1iq}(\tilde{\theta}) - \tilde{\boldsymbol{B}}_q^T \boldsymbol{z}_i$, $\tilde{\boldsymbol{B}}_q$ is obtained from $\tilde{\boldsymbol{B}}_p(\tilde{\theta})$ by changing $u_{ip}(\tilde{\theta})$ to $u_{2.1iq}(\tilde{\theta})$ and $u_{2.1iq}(\tilde{\theta})$ is the q-th element of $\boldsymbol{u}_{2.1i}(\tilde{\theta})$.

For stratified random sampling and stratified multistage sampling, we can also use a jackknife estimator, $\hat{V}_J(\tilde{\boldsymbol{S}}_2)$, or a jackknife linearization estimator of $V(\tilde{\boldsymbol{S}}_2)$. Following Rao, Scott and Skinner (1998) and the approach presented in subsections 4.2.2 and 4.3.2, it can be shown that the jackknife linearization estimator $\hat{V}_{JL}(\tilde{\boldsymbol{S}}_2)$ is identical to $\hat{V}(\tilde{\boldsymbol{S}}_2)$. Details are omitted for simplicity.

7. Computer Implementation

It is common practice for Statistical agencies to use auxiliary data in estimation procedures for a variety of sampling designs. In the mid-eighties, the need for

automated estimation systems incorporating the use of auxiliary data was recognized. In response to this need, several estimation software packages were developed; for example, LINWEIGHT (Bethlehem and Keller, 1987), PC-CARP (Shnell, Kennedy, Sullivan, Park and Fuller, 1988), SUDAAN (Shah, Lavange, Barnwell, Killinger, Wheeless, 1989), WESVAR (Brick and Morganstein, 1996), CLAN (Anderson and Nordberg, 2000), BASCULA (Nieuwenbreck and Hofman, 2000) and GES (Hidiroglou,. Estevao and Arcaro, 2000). All of these packages offer point and variance estimation for commonly used parameters such as totals, means and ratios, with some packages offering more than others. Differences between the packages include: (i) availability of analytic features such as linear regression, logistic regression and two-way table analysis; (ii) the methods of variance estimation (Taylor, jackknife or other replication methods). While the above packages estimate commonly used parameters, none of them seem to have the flexibility to handle arbitrary parameters of interest. On the other hand, the methods presented in this paper pave the way to automatically obtaining estimators and their corresponding standard errors through estimating equations. We now illustrate how the proposed methods can be implemented using Statistics Canada's Generalized Estimation System (GES).

Statistics Canada's GES system provides a modern framework for estimation and variance estimation. The framework adopted for building GES is based on the use of auxiliary information and of the calibration procedure of Deville and Särndal (1992). It is built around four key elements: (i) the sampling plan; (ii) the population parameters to be estimated; (iii) the use of auxiliary information, and (iv) the domains of interest. Several sampling designs incorporating with or without replacement sampling and probability proportional to size sampling are available. GES computes estimates, incorporating auxiliary data, of totals, means, and ratios with associated standard errors computed by Taylor linearization or jackknife procedures. The auxiliary information can cut across design strata, or be included within them. This allows the computation of most of the commonly used estimators in survey sampling, including separate and combined ratio or regression estimators (or intermediate combinations), poststratified estimators (separate, combined, or mixed) and others such as the raking ratio estimator. Estimates and their associated measures of reliability are computed for user-specified domains of interest. Further details of the methodology used in GES are given in Estevao, Hidiroglou and Särndal (1995).

GES was built as an integrated package and offers flexibility in terms of domain estimation and use of auxiliary data. However, it computes a limited number of estimates: means, totals, and ratios. The inclusion of additional parameters into GES is not straightforward as the existing software must be modified to incorporate new estimators within GES. This means that developers have to understand the architecture of GES, determine where the additional estimators fit in, program them, debug and document the resulting code.

A more flexible approach to implement new estimators into GES is to build the required software outside of GES and use existing components of GES where possible. For example, GES could be used to compute the estimation weights, a_i, the residuals, e_i, and the resulting standard errors within the existing variance estimation procedures in GES. We briefly describe the steps for integrating additional software with the existing components of GES.

1. Given the design weights, w_i, and the auxiliary data, z_i, GES can be used to compute the estimation weights a_i, given by (3.1), and the resulting calibration weights, $\tilde{w}_i = w_i a_i$.
2. An additional program is necessary to compute the estimator, $\hat{\boldsymbol{\theta}}$, of the parameter of interest, $\boldsymbol{\theta}_N$, using the estimating equation approach. The parameter of interest is defined by Census estimating equations of the form of (2.1). These estimating equations are completely specified by the $\boldsymbol{u}_i(\boldsymbol{\theta})$ terms which are defined by

$$u_{ip}(\boldsymbol{\theta}) = (\partial \mu_i / \partial \theta_p)\ \{(y_i - \mu_i)/V_{0i}\}.$$

   Thus, given the model mean and the model variance, the $\boldsymbol{u}_i(\boldsymbol{\theta})$ terms can be derived. Given the $\boldsymbol{u}_i(\boldsymbol{\theta})$ terms, the calibration weights, $\tilde{w}_i$, and the necessary data, we can then define the sample estimating equations as $\hat{\boldsymbol{S}}(\boldsymbol{\theta}) = \Sigma_s \tilde{w}_i \boldsymbol{u}_i(\boldsymbol{\theta})$. Finally, the program needs to apply the Newton-Raphson algorithm to solve the sample estimating equations to obtain $\hat{\boldsymbol{\theta}}$. As a byproduct to the Newton-Raphson method, the $\boldsymbol{J}(\hat{\boldsymbol{\theta}})$ matrix is also obtained. Note that for estimating equations with closed form solutions, the Newton-Raphson algorithm will converge to the correct solution in one iteration. Thus, for these estimating equations it is not necessary to explicitly define the closed form solutions, only the corresponding $\boldsymbol{u}_i(\boldsymbol{\theta})$ terms.
3. Next, given $\hat{\boldsymbol{\theta}}$, the $\boldsymbol{u}_i(\boldsymbol{\theta})$ terms evaluated at $\hat{\boldsymbol{\theta}}$ and the necessary data, GES can be used to compute the residuals, $\boldsymbol{e}_i^*$ given by (4.4).
4. Finally, a program is needed to calculate the synthetic residuals $\gamma_i = \boldsymbol{J}^{-1}(\hat{\boldsymbol{\theta}}) a_i \boldsymbol{e}_i^*$. GES can then be used to apply the appropriate variance operator to these synthetic residuals; that is,

use $\hat{V}(\hat{\theta}) = v(\gamma_i)$ to get the Taylor linearization estimate (or the jackknife linearization estimate where applicable) of $\hat{V}(\hat{\theta})$.

5. The jackknife estimators can also be implemented through the one-step jackknife, but details of the steps involved are not spelled out here.

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.

Andersson, C. and Nordberg, L. (2000). CLAN: A SAS-program for computation of point and standard error estimates in sample surveys. Invited paper to be presented at ICES 2, Buffalo, New York, June 19-21, 2000.

Bera, A.K. and Bilas, Y. (1998). Rao's Score, Nymen's $C(\alpha)$ and Silvey's LM tests: An essay on historical developments and some new results. Journal of Statistical Planning and Inference, in press.

Bethlehem, K.G. and Keller, W.K. (1987). Linear weighting of sample survey data. *Journal of Official Statistics*, **3**, 141-153.

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

Brick, J.M., and Morganstein, D.R. (1996). WesVarPC: Software for computing variance estimates from complex designs. *Proceedings of the Bureau of the Census 1996 Annual Research Conference.*

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, **11**, 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.

Hidiroglou, M.A., Estevao, V., and Arcaro, C. (2000). Generalised estimation system with future enhancement. Invited paper to be presented at ICES 2, Buffalo, New York, June 19-21, 2000.

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.

Nieuwenbreck, R. and Hofman L. (2000). Towards a generalized weighting system. Invited paper to be presented at ICES 2, Buffalo, New York, June 19-21, 2000.

Rao, C.R. (1947). Large sample tests of statistical hypothesis concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, **44**, 50-57.

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

Rao, J.N.K., Scott A.J. and Skinner, C.J. (1998). Quasi-score test with survey data. *Statistica Sinica*, **8**, 1059-1070.

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.

Shah, B.V., Lavange, L.M., Barnwell, B.G., Killinger, K.E., and Wheeless S.C. (1989). SUDAAN: Procedures for descriptive statistics user's guide. Research Triangle Institute Report.

Schnell, D., Kennedy, W.K., Sullivan, G., Park, K.P., and Fuller, W.A. (1988). Personal computer variance software for complex surveys. *Survey Methodology*, **14**, 59-69.

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