Sankhyā : The Indian Journal of Statistics
San Antonio Conference: Selected articles
2002, Volume 64, Series A, Pt.2, pp 364-378

ESTIMATING EQUATIONS FOR THE ANALYSIS OF SURVEY DATA USING POSTSTRATIFICATION INFORMATION

By J.N.K. RAO
Carleton University, Ottawa, Canada,
W. YUNG and M.A. HIDIROGLOU
Statistics Canada, Ottawa, Canada

SUMMARY. Finite population parameters may be expressed as solutions to suitable "census" estimating equations (EE). Parameter estimates are obtained by solving sample EE which involve design weights as well as adjustment factors (also called g-weights) based on poststratification information. Taylor linearization variance estimators that take account of poststratification adjustments are given and practical implications are noted. For stratified simple random sampling and stratified multistage sampling, jackknife linearization variance estimators are developed. Wald tests and analogues of C.R. Rao's score tests that take account of the survey design features and poststratification are also given. Finally, the implementation of the proposed methodology into a computer system is outlined.

1. **Introduction**

Many finite population parameters of interest, such as means, ratios, linear and logistic regression coefficients, and measures of income inequality, can be expressed as solutions to suitable "census" estimating equations (EE). Parameter estimates are obtained by solving sample EE which involve design weights as well as adjustment factors (also called g-weights) based on poststratification information.

Binder (1983), Godambe and Thompson (1986), Kovacevic and Binder (1997) and others studied the EE approach for making inference from complex survey data. They used the Taylor linearization method for variance

Paper received October 2000; revised December 2001.
AMS (1991) subject classification. 62D05, 62G09.
Keywords and phrases. Generalized regression estimator, jackknife, stratified multi-stage sampling.

estimation. However, the implication of poststratfication adjustment on variance estimation has not been fully studied in the literature. Commonly used software packages currently provide "proper" linearization variance estimators only for special cases, such as totals or ratios under one-way poststratification.

In Section 2, we provide proper linearization variance estimators for the general case of estimators obtained as solutions of sample EE, using one or more poststratification variables with known marginal totals. Practical implications pertaining to the efficiency of estimators are also noted. For stratified simple random sampling and stratified multistage sampling, jackknife variance estimators and jackknife linearization variance estimators, that take account of poststratification adjustments, are given in Section 3. Wald tests and analogues of C.R. Rao's (1948) score tests on model parameters that take account of survey design features and poststratification are presented in Section 4. Finally, Section 5 outlines the implementation of the proposed methodology in a computer system to provide a flexible and unified approach.

2. Linearization Variance Estimators

2.1. Parameters. A finite parameter vector $\boldsymbol{\theta}_N$ can be regarded as the solution to "census" estimation equations (EE)

$$\mathbf{S}(\boldsymbol{\theta}) = \sum_{k \in U} \mathbf{u}_k(\boldsymbol{\theta}) = \mathbf{0} \tag{2.1}$$

where $\sum_{k \in U}$ denotes the summation over the finite population U of size N, and the estimation functions $\mathbf{u}_k(\boldsymbol{\theta})$ are suitably chosen. For example, $\mathbf{u}_k(\boldsymbol{\theta}) = \mathbf{y}_k - \boldsymbol{\theta}$ in (2.1) gives the population mean $\boldsymbol{\theta}_N = \bar{\mathbf{Y}}$, $\mathbf{u}_k(\boldsymbol{\theta}) = \mathbf{x}_k(y_k - \mu_k(\boldsymbol{\theta}))$ with $\mu_k(\boldsymbol{\theta}) = \mathbf{x}_k^T \boldsymbol{\theta}$ gives the least squares regression vector $\boldsymbol{\theta}_N = \left(\sum_{k \in U} \mathbf{x}_k^T \mathbf{x}_k \right)^{-1} \sum_{k \in U} \mathbf{x}_k y_k$, and $\mathbf{u}_k(\boldsymbol{\theta}) = \mathbf{x}_k(y_k - \mu_k(\boldsymbol{\theta}))$ with $\mu_k(\boldsymbol{\theta}) = \exp(\mathbf{x}_k^T \boldsymbol{\theta}) \left[1 + \exp(\mathbf{x}_k^T \boldsymbol{\theta}) \right]^{-1}$ gives the logistic regression vector $\boldsymbol{\theta}_N$. Kovacevic and Binder (1997) provide estimating functions, $\mathbf{u}_k(\boldsymbol{\theta})$, that lead to various measures of income inequality, such as the Gini index and the polarization index.

Census parameters, $\boldsymbol{\theta}_N$, may be motivated through a model for the finite population. However, our interest, in Sections 2 and 3, is in making design-based inferences on the finite population parameters, $\boldsymbol{\theta}_N$, defined as the solution to census EE. In Section 4, we study tests on model parameters assuming a super-population model.

2.2. Estimation of $\boldsymbol{\theta}_N$. Poststratification is based on one or more poststratification variables. It is commonly used to improve the precision of estimators of a total $Y = \sum_{k\in U} y_k$, and to ensure benchmarking to known marginal totals corresponding to the categories of the poststratification variables. For example, projected age-sex census counts are used in socio-economic surveys.

Let $\mathbf{z}_k$ be the vector of indicator variables, associated with unit k, that correspond to the categories of the poststratification variables with known marginal totals $\mathbf{Z} = \sum_{k\in U} \mathbf{z}_k$. Benchmarking is accomplished by adjusting the design weights, w_k for $k \in s$, to construct calibration weights $\tilde{w}_k(s)$ that satisfy $\sum_{k\in s} \tilde{w}_k(s)\mathbf{z}_k = \mathbf{Z}$, where $\sum_{k\in s}$ denotes the summation over the sample s. The calibration weight $\tilde{w}_k(s)$ may be expressed as the product of the design weight w_k and a poststratification adjustment factor $a_k(s)$. Generalized regression (GREG) is commonly used for constructing calibration weights. In this case, $a_k(s)$ is given by

$$a_k(s) = \mathbf{z}_k^T \left(\sum_{k\in s} w_k \mathbf{z}_k \mathbf{z}_k^T \right)^{-1} \mathbf{Z}. \tag{2.2}$$

The adjustment factor (2.2) is also called the g-weight (see Särndal, Swensson and Wretman, 1992, chapter 6). The GREG estimator of a vector of totals, $\mathbf{Y} = \sum_{k\in U} \mathbf{y}_k$, is given by

$$\hat{\mathbf{Y}} = \sum_{k\in s} \tilde{w}_k(s)\mathbf{y}_k,$$

where $\tilde{w}_k(s) = w_k a_k(s)$ and $a_k(s)$ is given by (2.2). We focus on GREG calibration weights in this paper.

The GREG estimator, $\hat{\boldsymbol{\theta}}$, of $\boldsymbol{\theta}_N$ is obtained as the solution to the sample EE given by

$$\hat{\mathbf{S}}(\boldsymbol{\theta}) = \sum_{k\in s} \tilde{w}_k(s)\mathbf{u}_k(\boldsymbol{\theta}) = \mathbf{0}, \tag{2.3}$$

noting that $\mathbf{S}(\boldsymbol{\theta})$ is the total of the $\mathbf{u}_k(\boldsymbol{\theta})$'s. The solution to (2.3) has a closed form in simple cases (e.g., mean, least squares regression). For logistic regression and other complex cases, it is necessary to solve (2.3) iteratively to obtain the solution $\hat{\boldsymbol{\theta}}$. The Newton-Raphson (N-R) algorithm is commonly used for solving (2.3). The r-th step of the N-R algorithm is given by

$$\hat{\boldsymbol{\theta}}_r = \hat{\boldsymbol{\theta}}_{r-1} + \hat{\mathbf{J}}^{-1}(\hat{\boldsymbol{\theta}}_{r-1})\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}}_{r-1}), \tag{2.4}$$

where $\hat{\boldsymbol{\theta}}_{r-1}$ is the value of $\hat{\boldsymbol{\theta}}$ obtained at the (r-1)-th iteration and $\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}}_{r-1})$ and $\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}}_{r-1})$ are the values of $\hat{\mathbf{J}}(\boldsymbol{\theta}) = -\partial\hat{\mathbf{S}}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^T$ and $\hat{\mathbf{S}}(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta} =$

$\hat{\boldsymbol{\theta}}_{r-1}$. Iterating the algorithm to convergence produces the GREG estimator $\hat{\boldsymbol{\theta}}$.

2.3. Variance Estimation. Let $\mathbf{v}(\mathbf{y}_k)$ denote the estimator of the covariance matrix of the basic design-unbiased estimator $\hat{\mathbf{Y}}_b = \sum_{k\in s} w_k \mathbf{y}_k$. Then a linearization variance estimator of $\hat{\mathbf{Y}}$ is given by

$$\hat{\mathbf{V}}_L(\hat{\mathbf{Y}}) = \mathbf{v}(a_k(s)\mathbf{e}_k), \tag{2.5}$$

where the j-th element of $\mathbf{e}_k$ is given by $e_{kj} = y_{kj} - \hat{\mathbf{B}}_j^T \mathbf{z}_k$ and $\hat{\mathbf{B}}_j = \left(\sum_{k\in s} w_k \mathbf{z}_k \mathbf{z}_k^T\right)^{-1} \sum_{k\in s} w_k \mathbf{z}_k y_{kj}$; see Särndal, Swensson and Wretman (1989). An alternative linearization variance estimator of $\hat{\mathbf{Y}}$ is given by $v(\mathbf{e}_k)$. In the context of the ratio estimator $\hat{Y}_r = (\bar{y}/\bar{x})X$ under simple random sampling, Royall and Cumberland (1981) showed that (2.5) tracks the conditional variance of $\hat{Y}_r$ given the sample mean $\bar{x}$ better than the alternative variance estimator. Valliant (1993) conducted a simulation study for one-way poststratification and demonstrated that (2.5) possesses good conditional properties given the estimated poststrata counts. Särndal, Swensson and Wretman (1989) showed that (2.5) is a good choice from either the design-based or the model-based perspective.

Under regularity conditions, Binder (1983) obtained a linearization variance estimator of $\hat{\boldsymbol{\theta}}$ as

$$\hat{\mathbf{V}}_L(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{J}}^{-1}(\hat{\boldsymbol{\theta}})\hat{\mathbf{V}}\left(\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})\right)\hat{\mathbf{J}}^{-1}(\hat{\boldsymbol{\theta}}), \tag{2.6}$$

where $\hat{\mathbf{V}}(\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}}))$ is a variance estimator of $\hat{\mathbf{S}}(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. However, he did not study the implications of poststratification on the variance estimator of $\hat{\mathbf{S}}(\boldsymbol{\theta})$. Noting that $\hat{\mathbf{S}}(\boldsymbol{\theta})$ is the GREG estimator of the total $\mathbf{S}(\boldsymbol{\theta}) = \sum_{k\in U} \mathbf{u}_k(\boldsymbol{\theta})$, we can appeal to (2.5) to get a variance estimator of $\hat{\mathbf{S}}(\boldsymbol{\theta})$. We have (Hidiroglou, Rao and Yung, 1999)

$$\hat{\mathbf{V}}(\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})) = \mathbf{v}[a_k(s)\mathbf{e}_k^*(\hat{\boldsymbol{\theta}})], \tag{2.7}$$

where the j-th element of $\mathbf{e}_k^*(\hat{\boldsymbol{\theta}})$ is given by

$$e_{kj}^*(\hat{\boldsymbol{\theta}}) = u_{kj}(\hat{\boldsymbol{\theta}}) - \hat{\mathbf{B}}_j^T(\hat{\boldsymbol{\theta}})\mathbf{z}_k, \tag{2.8}$$

and $\hat{\mathbf{B}}_j(\hat{\boldsymbol{\theta}})$ is obtained from $\hat{\mathbf{B}}_j$ by changing y_{kj} to $u_{kj}(\hat{\boldsymbol{\theta}})$:

$$\hat{\mathbf{B}}_j(\hat{\boldsymbol{\theta}}) = \left(\sum_{k\in s} w_k \mathbf{z}_k \mathbf{z}_k^T\right)^{-1} \sum_{k\in s} w_k \mathbf{z}_k u_{kj}(\hat{\boldsymbol{\theta}}). \tag{2.9}$$

It follows from (2.6) and (2.7) that $\hat{\mathbf{V}}_L(\hat{\boldsymbol{\theta}})$ may be expressed as

$$\hat{\mathbf{V}}_L(\hat{\boldsymbol{\theta}}) = \mathbf{v}\left[a_k(s)\tilde{\mathbf{e}}_k(\hat{\boldsymbol{\theta}})\right], \tag{2.10}$$

where $\tilde{\mathbf{e}}_k(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{J}}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{e}_k^*(\hat{\boldsymbol{\theta}})$. It follows from (2.10) that the linearization variance estimator, $\hat{\mathbf{V}}_L(\hat{\boldsymbol{\theta}})$, is obtained from the formula $\mathbf{v}(\mathbf{y}_k)$ for the basic estimator $\hat{\mathbf{Y}}_b$ by changing $\mathbf{y}_k$ to $a_k(s)\tilde{\mathbf{e}}_k(\hat{\boldsymbol{\theta}})$.

An important practical implication of (2.10) in the context of linear or logistic regression is that the poststratification may not lead to a significant gain in efficiency if the "residuals" $u_{kj}(\boldsymbol{\theta})$ are unrelated to the poststratification variables. This can happen when the underlying model provides a good fit to the data. On the other hand, poststratification can lead to significant gains in efficiency in the context of estimating the total Y. For example, consider the case of one-way poststratification with estimated poststrata means $\hat{\bar{Y}}_j$. In this case $e_{kj}^* = y_k - \hat{\bar{Y}}_j$ if unit k belongs to poststratum j and the deviations e_{kj}^* will be less variable within poststrata compared to the deviations $y_k - \hat{\bar{Y}}_b$ associated with the variance estimator $v(y_k)$ of the basic estimator $\hat{Y}_b$, where $\hat{\bar{Y}}_b$ is the basic estimator of the mean $\bar{Y} = Y/N$. Silva (1996; p. 133) and Deville (1999; section 10) mentioned a variance estimator of the form (2.10) without discussing its practical implications in the context of linear or logistic regression.

Example 2.1. Stratified simple random sampling is commonly used in establishment surveys based on list frames. Suppose n_h units are selected by simple random sampling from N_h units in the h-th stratum, $h = 1, \ldots, L$, independently across strata. The basic weight attached to the i-th sample unit in the h-th stratum is given by $w_{hi} = N_h/n_h, i = 1, \ldots, n_h$ and the associated g-weight, $a_{hi}(s)$, is given by (2.2) with the subscript k changed to hi. The estimated covariance matrix of $\hat{\mathbf{Y}}_b = \sum_{(hi)\in s} w_{hi}\mathbf{y}_{hi}$ is given by

$$\hat{\mathbf{V}}(\hat{\mathbf{Y}}_b) = \mathbf{v}(\mathbf{y}_{hi}) = \sum_h N_h^2(1-f_h)\frac{1}{n_h(n_h-1)}\sum_i(\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)(\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)^T, \tag{2.11}$$

where $f_h = n_h/N_h$ and $\bar{\mathbf{y}}_h = \sum_i \mathbf{y}_{hi}/n_h$. It follows from (2.10) that

$$\hat{\mathbf{V}}_L(\hat{\boldsymbol{\theta}}) = \mathbf{v}\left[a_{hi}(s)\tilde{\mathbf{e}}_{hi}(\hat{\boldsymbol{\theta}})\right], \tag{2.12}$$

which is obtained from (2.11) by changing $\mathbf{y}_{hi}$ to $a_{hi}(s)\tilde{\mathbf{e}}_{hi}(\hat{\boldsymbol{\theta}})$.

Example 2.2. Stratified multi-stage sampling is commonly used in large socio-economic surveys. We focus on designs with a large number of strata,

L, and relatively few clusters, $n_h(\geq 2)$, sampled within each stratum $h(= 1,\ldots,L)$. We assume that subsampling within sampled clusters $i(= 1,\ldots,n_h)$ is performed to ensure design-unbiased estimation of cluster totals. Denote the design weight attached to the k-th sample unit in the (hi)-th sample cluster as w_{hik} and the associated g-weight as $a_{hik}(s)$. For variance estimation, the clusters are assumed to be sampled with replacement. The estimated covariance matrix of the basic estimator $\hat{\mathbf{Y}}_b = \sum_{(hik)\in s} w_{hik}\mathbf{y}_{hik}$ is then given by

$$\hat{\mathbf{V}}(\hat{\mathbf{Y}}_b) = \mathbf{v}(\mathbf{y}_{hi}) = \sum_h [n_h(n_h-1)]^{-1}\sum_i(\mathbf{y}_{hi}-\bar{\mathbf{y}}_h)(\mathbf{y}_{hi}-\bar{\mathbf{y}}_h)^T, \qquad (2.13)$$

where $\mathbf{y}_{hi} = \sum_k(n_h w_{hik})\mathbf{y}_{hik}$ and $\bar{\mathbf{y}}_h = \sum_i \mathbf{y}_{hi}/n_h$. Note that $\hat{\mathbf{V}}(\hat{\mathbf{Y}}_b)$ depends only on the cluster totals, $\mathbf{y}_{hi}$. The variance estimator (2.13) leads to overestimation, but the relative bias is likely to be small if the first-stage sampling fractions are small. It follows from (2.10) and (2.13) that

$$\hat{\mathbf{V}}_L(\hat{\boldsymbol{\theta}}) = \mathbf{v}\left[a_{hik}(s)\tilde{\mathbf{e}}_{hi}(\hat{\boldsymbol{\theta}})\right], \qquad (2.14)$$

where $\tilde{\mathbf{e}}_{hi}(\hat{\boldsymbol{\theta}}) = \sum_k(n_h w_{hik})\tilde{\mathbf{e}}_{hik}(\hat{\boldsymbol{\theta}})$.

2.4. Confidence Intervals. Suppose we are interested in confidence intervals on the components of $\boldsymbol{\theta}$; for example, the first component θ_1. Then we obtain the variance estimator of $\hat{\theta}_1$ from (2.10) as $V_L(\hat{\theta}_1) = v[a_k(s)\tilde{e}_{k1}(\hat{\boldsymbol{\theta}})]$, where $\tilde{e}_{k1}(\hat{\boldsymbol{\theta}})$ is the first element of $\tilde{\mathbf{e}}_k(\hat{\boldsymbol{\theta}})$. A $(1-\alpha)$-level normal theory confidence interval on θ_1 is given by $[\hat{\theta}_1 - z_{\alpha/2}v^{1/2}(a_k(s)\tilde{e}_{k1}(\hat{\boldsymbol{\theta}})), \hat{\theta}_1 + z_{\alpha/2}v^{1/2}(a_k(s)\tilde{e}_{k1}(\hat{\boldsymbol{\theta}}))]$, where $z_{\alpha/2}$ is the upper $\alpha/2$-point of a $N(0,1)$ variable. This interval is asymptotically correct under regularity conditions.

Denote $\mathbf{u}_k(\boldsymbol{\theta}) = [u_{k1}(\boldsymbol{\theta}), \mathbf{u}_{k2}(\boldsymbol{\theta})^T]^T$ and suppose that $\mathbf{u}_{k2}(\boldsymbol{\theta})$ does not depend on θ_1. Also, suppose that the first derivative of $u_{k1}(\boldsymbol{\theta})$ with respect to θ_1 is independent of $\boldsymbol{\theta}_2$, where $\boldsymbol{\theta} = (\theta_1, \boldsymbol{\theta}_2^T)^T$. Then, expressing $\hat{\mathbf{J}}(\boldsymbol{\theta})$ as a partitioned matrix with diagonals $\{\hat{J}_{11}(\boldsymbol{\theta}), \hat{\mathbf{J}}_{22}(\boldsymbol{\theta})\}$ and off diagonals as $\{\hat{\mathbf{J}}_{12}(\boldsymbol{\theta}), \hat{\mathbf{J}}_{21}(\boldsymbol{\theta})\}$, we have $\hat{\mathbf{J}}_{12}(\boldsymbol{\theta}) = \mathbf{0}$, $\hat{J}_{11}(\boldsymbol{\theta}) = \hat{J}_{11}(\theta_1)$ and $\tilde{e}_{k1}(\hat{\boldsymbol{\theta}})$ reduces to

$$\tilde{e}_{k1}(\hat{\boldsymbol{\theta}}) = J_{11}^{-1}(\hat{\theta}_1)\left[\hat{\mathbf{J}}_{12}(\hat{\boldsymbol{\theta}})\hat{\mathbf{J}}_{22}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{e}_{k2}^*(\hat{\boldsymbol{\theta}}) - e_{k1}^*(\hat{\boldsymbol{\theta}})\right]. \qquad (2.15)$$

Kovacevic and Binder (1997) studied the above special case in the context of measures of income inequality, but poststratification was not taken into account for variance estimation. That is, $\mathbf{u}_k(\hat{\boldsymbol{\theta}})$ was used in (2.15) instead of $\mathbf{e}_k^*(\hat{\boldsymbol{\theta}})$.

3. Jackknife Variance Estimators

The Taylor linearization approach works for general designs, whereas the jackknife method of variance estimation is restricted to particular designs. The jackknife is applicable for the two commonly used designs given in Examples 2.1 and 2.2. We present jackknife variance estimators and derive a jackknife linearization variance estimator by approximating the jackknife.

Example 2.1 (contd.). For the stratified random sampling design, we define the jackknife weights when the j-th sample unit in the l-th stratum is deleted as $w_{hi(lj)} = 0$ if $(hi) = (lj)$; $= n_l/(n_l-1)w_{li}$ if $h = l, i \neq j$; $= w_{hi}$ if $h \neq l$. The jackknife calibration weights, $\tilde{w}_{hi(lj)}(s)$, are obtained from the original calibration weights, $\tilde{w}_{hi}(s)$, by changing w_{hi} to $w_{hi(lj)}$. The resulting sample EE is given by

$$\hat{\mathbf{S}}_{(lj)}(\boldsymbol{\theta}) = \sum_{(hi)\in s} \tilde{w}_{hi(lj)}(s)\mathbf{u}_{hi}(\boldsymbol{\theta}) = \mathbf{0}. \tag{3.1}$$

To obtain the solution, $\hat{\boldsymbol{\theta}}_{(lj)}$, from (3.1), we can use the Newton-Raphson (N-R) algorithm (2.4) until convergence, but it is computer intensive. Alternatively, a one-step N-R with $\hat{\boldsymbol{\theta}}$ as the starting value is computationally simpler than the full N-R. The one-step N-R leads to

$$\hat{\boldsymbol{\theta}}_{(lj)} = \hat{\boldsymbol{\theta}} + \hat{\mathbf{J}}^{-1}_{(lj)}(\hat{\boldsymbol{\theta}})\hat{\mathbf{S}}_{(lj)}(\hat{\boldsymbol{\theta}}), \tag{3.2}$$

where $\hat{\mathbf{J}}_{(lj)}(\hat{\boldsymbol{\theta}})$ is obtained from $\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}})$ by changing $\tilde{w}_{hi}(s)$ to $\tilde{w}_{hi(lj)}(s)$.

A one-step jackknife variance estimator of $\hat{\boldsymbol{\theta}}$ is given by

$$\hat{\mathbf{V}}_J(\hat{\boldsymbol{\theta}}) = \sum_l (1-f_l)\frac{n_l-1}{n_l}\sum_j \left(\hat{\boldsymbol{\theta}}_{(lj)} - \hat{\boldsymbol{\theta}}\right)\left(\hat{\boldsymbol{\theta}}_{(lj)} - \hat{\boldsymbol{\theta}}\right)^T. \tag{3.3}$$

To obtain a jackknife linearization variance estimator of $\hat{\boldsymbol{\theta}}$, we note that $\hat{\mathbf{J}}^{-1}_{(lj)}(\hat{\boldsymbol{\theta}}) \approx \hat{\mathbf{J}}^{-1}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ so that (3.2) may be approximated as

$$\hat{\boldsymbol{\theta}}_{(lj)} - \hat{\boldsymbol{\theta}} \approx \hat{\mathbf{J}}^{-1}(\hat{\boldsymbol{\theta}})\left[\hat{\mathbf{S}}_{(lj)}(\hat{\boldsymbol{\theta}}) - \hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})\right]. \tag{3.4}$$

It now follows from (3.3) and (3.4) that

$$\hat{\mathbf{V}}_J(\hat{\boldsymbol{\theta}}) \approx \hat{\mathbf{J}}^{-1}(\hat{\boldsymbol{\theta}})\hat{\mathbf{V}}_J\left[\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})\right]\hat{\mathbf{J}}^{-1}(\hat{\boldsymbol{\theta}}), \tag{3.5}$$

where $\hat{\mathbf{V}}_J\left[\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})\right]$ is the jackknife variance estimator of $\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})$. Further, noting that $\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})$ is a GREG estimator of the total $\mathbf{S}(\boldsymbol{\theta})$, we can approximate $\hat{\mathbf{V}}_J\left[\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})\right]$ by a jackknife linearization variance estimator given by

$$\hat{\mathbf{V}}_{JL}\left[\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})\right] = \mathbf{v}\left[a_{hi}(s)\mathbf{e}^*_{hi}(\hat{\boldsymbol{\theta}})\right]. \tag{3.6}$$

The result (3.6) can be obtained by applying the method of Yung and Rao (1996) who considered the GREG estimation of the total Y. It now follows from (3.5) and (3.6) that

$$\hat{\mathbf{V}}_{JL}(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{V}}_L(\hat{\boldsymbol{\theta}}), \tag{3.7}$$

where $\hat{\mathbf{V}}_L(\hat{\boldsymbol{\theta}})$ is given by (2.12). Hence, jackknife linearization is identical to Taylor linearization that uses the g-weighted residuals.

Example 2.2 (contd.). For the stratified multistage sampling design, the jackknife weights $w_{hik(lj)}$ when the (lj)-th sample cluster is deleted are given by $w_{hik(lj)} = 0$ if $(hi) = (lj)$; $= n_l/(n_l - 1)w_{lik}$ is $h = l, i \neq j$; $= w_{hik}$ if $h \neq l$. We calculate $\hat{\boldsymbol{\theta}}_{(lj)}$ as in the case of Example 2.1, using the one-step N-R method. A one-step jackknife variance estimator of $\hat{\boldsymbol{\theta}}$ is given by

$$\hat{\mathbf{V}}_J(\hat{\boldsymbol{\theta}}) = \sum_l \frac{n_h - 1}{n_h} \sum_j \left(\hat{\boldsymbol{\theta}}_{(lj)} - \hat{\boldsymbol{\theta}}\right)\left(\hat{\boldsymbol{\theta}}_{(lj)} - \hat{\boldsymbol{\theta}}\right)^T. \tag{3.8}$$

Now, following along the lines of Example 2.1, we get a jackknife linearization variance estimator from (3.8) which is identical to the Taylor linearization variance estimator (2.14).

4. Wald Tests and Quasi-Score Tests

In Sections 2 and 3 we focussed on design-based inferences pertaining to a census parameter vector $\boldsymbol{\theta}_N$ defined as the solution to a suitably defined census EE. In this section, we turn to design-model inferences on a model parameter vector $\boldsymbol{\theta}$, assuming a super-population model. In particular, we are interested in testing a null hypothesis of the form $H_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{20}$, where $\boldsymbol{\theta}$ is partitioned as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ with $\boldsymbol{\theta}_2$ a $q \times 1$ vector and $\boldsymbol{\theta}_1$ a $r \times 1$ vector. We assume that the census y-values are generated by a random process with mean $E_m(y_k) = \mu_k = \mu(\mathbf{x}_k, \boldsymbol{\theta})$ and "working" variance $V_m(y_k) = V_{0k} = \sigma^2 V_0(\mu_k)$ for some known $V(\cdot)$, where E_m and V_m denote model expectation and model variance, respectively. The model for the mean is assumed to be valid, but the working variance may only be a rough approximation at best. The $\mathbf{u}$-function is given by $\mathbf{u}_k(\boldsymbol{\theta}) = (u_{k1}(\boldsymbol{\theta}), \ldots, u_{kp}(\boldsymbol{\theta}))^T$ with

$$u_{kl}(\boldsymbol{\theta}) = (\partial\mu_k/\partial\theta_l)(y_k - \mu_k)/V_{0k},\ l = 1, \ldots, p. \tag{4.1}$$

Substituting the above $\mathbf{u}_k(\boldsymbol{\theta})$ in (2.1) we get the census EE and the census parameter $\boldsymbol{\theta}_N$. Similarly, the GREG estimator, $\hat{\boldsymbol{\theta}}$, is obtained from (2.3).

4.1. Wald Tests. One approach to testing H_0 is to use the Wald statistic, W, based on $\hat{\boldsymbol{\theta}}$ and the variance estimator $\hat{\mathbf{V}}_L(\hat{\boldsymbol{\theta}})$ given by (2.10):

$$W = (\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20})^T \hat{\mathbf{V}}_{L22}^{-1} (\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20}), \tag{4.2}$$

where $\hat{\mathbf{V}}_{L22}$ is the lower diagonal block of the matrix $\hat{\mathbf{V}}_L(\hat{\boldsymbol{\theta}})$, given by (2.10), partitioned in correspondence with $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1^T, \hat{\boldsymbol{\theta}}_2^T)^T$. For stratified random sampling or stratifed multi-stage sampling, we can use the jackknife estimator $\hat{\mathbf{V}}_{J22}$.

We write $n_e^{1/2}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20})$ as

$$\hat{\mathbf{T}} = n_e^{1/2}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20}) = n_e^{1/2}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{2N}) + (n_e/N_e)^{1/2} N_e^{1/2}(\boldsymbol{\theta}_{2N} - \boldsymbol{\theta}_{20}), \tag{4.3}$$

where n_e and N_e are the "effective" sample and population sizes respectively and $\boldsymbol{\theta}_N = (\boldsymbol{\theta}_{1N}^T, \boldsymbol{\theta}_{2N}^T)^T$ is the census parameter. For example, under two-stage cluster sampling, we may take n_e and N_e as the number of sample clusters and the number of population clusters respectively, assuming that the population is generated by a model that induces correlated model errors within clusters.

Under regularity conditions, $\hat{\mathbf{T}}_1 = n_e^{1/2}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{2N})$ converges in design distribution to $N_q(\mathbf{0}, \boldsymbol{\Sigma}_{22N})$, say and $n_e\hat{\mathbf{V}}_{L22}$ converges in design probability to $\boldsymbol{\Sigma}_{22N}$. Similarly, $\mathbf{T}_{2N} = N_e^{1/2}(\hat{\boldsymbol{\theta}}_{2N} - \boldsymbol{\theta}_{20})$ converges in model distribution, under H_0, to $N_q(\mathbf{0}, \boldsymbol{\Sigma}_{22}^*)$, say. Rubin-Bleuer and Schiopu-Kratina (2001) have shown that $\hat{\mathbf{T}}_1$ and $\mathbf{T}_{2N}$ are asymptotically independent so that $\hat{\mathbf{T}}$, given by (4.3), converges in distribution to $N_q(\mathbf{0}, \boldsymbol{\Sigma}_{22} + f_e\boldsymbol{\Sigma}_{22}^*)$, where $f_e = \lim(n_e/N_e)$ and $\boldsymbol{\Sigma}_{22N}$ converges in model probability to $\boldsymbol{\Sigma}_{22}$. If $f_e = 0$, then it follows that the Wald statistic W, given by (4.2), is asymptotically χ_q^2 under H_0. Otherwise, it is necessary to replace $n_e\hat{\mathbf{V}}_{L22}$ by a consistent (design-model) estimator of $\boldsymbol{\Sigma}_{22} + f_e\boldsymbol{\Sigma}_{22}^*$. This would require the specification of the error covariance structure of the super-population model (Rubin-Bleuer and Schiopu-Kratina, 2001; Korn and Graubard, 1998). We follow Pfeffermann (1993) and Skinner, Holt and Smith (1989, p. 141) and assume that $f_e = 0$ which is typically satisfied in large scale surveys if the model errors associated with the finite population are regarded as independent. Under this assumption, we can use the observed value of W, say $W(obs)$, to calculate a p-value as $Pr[\chi_q^2 \geq W(obs)]$ under the design-model distribution.

4.2. Quasi-Score Tests. Wald tests have the following drawbacks; (i) The full model needs to be fitted to get $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1^T, \hat{\boldsymbol{\theta}}_2^T)^T$. (ii) Wald tests are not invariant to nonlinear transformations of $\boldsymbol{\theta}$ (Boos, 1992). C.R. Rao's (1948) score tests are free of both limitations and permit extensions to complex problems (as in the case of Wald tests). We need only fit the simpler null model, which is a considerable advantage if the full model contains a large number of parameters as in the case, for example, with a factorial structure containing a large number of interaction effects, $\boldsymbol{\theta}_2$, and the null hypothesis is the absence of interactions, i.e., $H_0 : \boldsymbol{\theta}_2 = \mathbf{0}$. A special issue of the Journal of Statistical Planning and Inference (2001) is devoted to recent developments related to Rao's score tests.

Rao, Scott and Skinner (1998) extended Rao's score tests to survey data by developing quasi-score tests that take account of survey design features. However, the effects of calibration were not studied. In this section, we give a brief account of quasi-score tests when calibration weights, $\tilde{w}_k(s)$, are used. Let $\hat{\mathbf{S}}(\boldsymbol{\theta}) = (\hat{\mathbf{S}}_1(\boldsymbol{\theta})^T, \hat{\mathbf{S}}_2(\boldsymbol{\theta})^T)^T$ be the partition of $\hat{\mathbf{S}}(\boldsymbol{\theta})$ corresponding to the partition of $\boldsymbol{\theta}$, where $\hat{\mathbf{S}}_t(\boldsymbol{\theta}) = \sum_{k \in s} \tilde{w}_k(s)\mathbf{u}_{tk}(\boldsymbol{\theta})$, $t = 1, 2$ and $\mathbf{u}_k(\boldsymbol{\theta}) = (\mathbf{u}_{1k}(\boldsymbol{\theta})^T, \mathbf{u}_{2k}(\boldsymbol{\theta})^T)^T$. Also, let $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\theta}}_1^T, \tilde{\boldsymbol{\theta}}_{20}^T)^T$ be the solution of $\hat{\mathbf{S}}_1(\boldsymbol{\theta}) = \mathbf{0}$ when $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_{20}$. The analogue of Rao's score test of $H_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{20}$, which we call the quasi-score test, is based on the statistic

$$QS = \tilde{\mathbf{S}}_2^T[\hat{\mathbf{V}}(\tilde{\mathbf{S}}_2)]^{-1}\tilde{\mathbf{S}}_2 \tag{4.4}$$

where $\tilde{\mathbf{S}}_2 = \hat{\mathbf{S}}_2(\tilde{\boldsymbol{\theta}})$ and $\hat{\mathbf{V}}(\tilde{\mathbf{S}}_2)$ is a consistent estimator of $\mathbf{V}(\tilde{\mathbf{S}}_2)$. We expand $\hat{\mathbf{S}}_1(\tilde{\boldsymbol{\theta}})$ and $\hat{\mathbf{S}}_2(\tilde{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_{20}^T)^T$ to get

$$\mathbf{0} = \hat{\mathbf{S}}_1(\tilde{\boldsymbol{\theta}}) = \hat{\mathbf{S}}_1(\boldsymbol{\theta}^*) - \hat{\mathbf{J}}_{11}(\boldsymbol{\theta}^*)(\tilde{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) \tag{4.5}$$

and

$$\hat{\mathbf{S}}_2(\tilde{\boldsymbol{\theta}}) \approx \hat{\mathbf{S}}_2(\boldsymbol{\theta}^*) - \hat{\mathbf{J}}_{21}(\boldsymbol{\theta}^*)(\tilde{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1), \tag{4.6}$$

where

$$\hat{\mathbf{J}}(\boldsymbol{\theta}^*) = \begin{bmatrix} \hat{\mathbf{J}}_{11}(\boldsymbol{\theta}^*) & \hat{\mathbf{J}}_{12}(\boldsymbol{\theta}^*) \\ \hat{\mathbf{J}}_{21}(\boldsymbol{\theta}^*) & \hat{\mathbf{J}}_{22}(\boldsymbol{\theta}^*) \end{bmatrix}.$$

The approximations (4.5) and (4.6) are justified because $\tilde{\boldsymbol{\theta}}_1$ is design-model consistent for $\boldsymbol{\theta}_1$ under H_0 (see Section 4.1). We now replace $\hat{\mathbf{J}}(\boldsymbol{\theta}^*)$ by its design-model expected value $\mathbf{I}(\boldsymbol{\theta}^*)$, say, in (4.5) and (4.6) and substitute for $\tilde{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1$ from (4.5) into (4.6) to get

$$\begin{aligned} \tilde{\mathbf{S}}_2 &= \hat{\mathbf{S}}_2(\tilde{\boldsymbol{\theta}}) \approx \hat{\mathbf{S}}_2(\boldsymbol{\theta}^*) - \mathbf{A}(\boldsymbol{\theta}^*)\hat{\mathbf{S}}_1(\boldsymbol{\theta}^*) \\ &= \sum_{k\in s} \tilde{w}_k(s)\tilde{\mathbf{u}}_{2k}(\boldsymbol{\theta}^*) \end{aligned} \tag{4.7}$$

where

$$\tilde{\mathbf{u}}_{2k}(\boldsymbol{\theta}^*) = \mathbf{u}_{2k}(\boldsymbol{\theta}^*) - \mathbf{A}(\boldsymbol{\theta}^*)\mathbf{u}_{1k}(\boldsymbol{\theta}^*)$$

and

$$\mathbf{A}(\boldsymbol{\theta}^*) = \mathbf{I}_{21}(\boldsymbol{\theta}^*)\mathbf{I}_{11}^{-1}(\boldsymbol{\theta}^*).$$

Using (4.7), we may write

$$\begin{aligned} \tilde{\mathbf{T}} = (n_e^{1/2}/N)\tilde{\mathbf{S}}_2 &\approx (n_e^{1/2}/N)\left[\sum_{k\in s} \tilde{w}_k(s)\tilde{\mathbf{u}}_{2k}(\boldsymbol{\theta}^*) - \sum_{k\in U} \tilde{\mathbf{u}}_{2k}(\boldsymbol{\theta}^*)\right] \\ &\quad +(n_e/N_e)^{1/2}(N_e^{1/2}/N)\sum_{k\in U} \tilde{\mathbf{u}}_{2k}(\boldsymbol{\theta}^*) \\ &= \tilde{\mathbf{T}}_1 + (n_e/N_e)^{1/2}\tilde{\mathbf{T}}_{2N}, \end{aligned} \tag{4.8}$$

where $\tilde{\mathbf{T}}_{2N}$ converges in model distribution to $N_q(\mathbf{0}, \tilde{\boldsymbol{\Sigma}}_{22}^*)$, say, and $\tilde{\mathbf{T}}_1$ converges in design distribution to $N_q(\mathbf{0}, \tilde{\boldsymbol{\Sigma}}_{22N})$, say. If $f_e \neq 0$ then it is necessary to specify the error covariance structure of the model to obtain a consistent estimator of $\tilde{\boldsymbol{\Sigma}}_{22}^*$. We assume that $f_e = 0$ here, as in the case of the Wald test. Then, $n_e\hat{\mathbf{V}}(\tilde{\mathbf{S}}_2)$ is a consistent estimator of $\tilde{\boldsymbol{\Sigma}}_{22N}$ and QS is asymptotically χ_q^2 under H_0. Under the assumption $f_e = 0$, we can use the observed value of QS, say $QS(obs)$, to calculate a p-value as $Pr[\chi_q^2 \geq QS(obs)]$ under the design-model distribution.

Noting that $\sum_{k\in s} \tilde{w}_k(s)\tilde{\mathbf{u}}_{2k}(\boldsymbol{\theta}^*)$ is a GREG estimator of the total $\sum_{k\in U} \tilde{\mathbf{u}}_{2k}(\boldsymbol{\theta}^*)$, we can use the linearization variance estimator $\hat{\mathbf{V}}_L(\tilde{\mathbf{S}}_2)$, obtained from (2.5), with y_{kj} replaced by the j-th element of $\mathbf{u}_{2k}(\tilde{\boldsymbol{\theta}}) - \mathbf{A}(\tilde{\boldsymbol{\theta}})\mathbf{u}_{1k}(\tilde{\boldsymbol{\theta}})$, where $\mathbf{A}(\tilde{\boldsymbol{\theta}})$ is an estimator of $\mathbf{A}(\boldsymbol{\theta}^*)$. A natural estimator of $\mathbf{A}(\boldsymbol{\theta}^*)$ is based on $\hat{\mathbf{J}}(\tilde{\boldsymbol{\theta}})$, but the resulting quasi-score test does not have the desired invariance property in general (Boos, 1992). We can get an invariant test by taking the expectation of $\hat{\mathbf{J}}(\boldsymbol{\theta}^*)$ under the mean specification $E_m(y_k) = \mu(\mathbf{x}_k, \boldsymbol{\theta})$. This leads to

$$\tilde{\mathbf{I}}(\tilde{\boldsymbol{\theta}}) = \sum_{k\in s} \tilde{w}_k(s)d_k(\tilde{\boldsymbol{\theta}})d_k(\tilde{\boldsymbol{\theta}})^T/V_{0k}(\tilde{\mu}_k), \tag{4.9}$$

where $d_k(\boldsymbol{\theta}) = \partial\mu(\mathbf{x}_k, \boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $\tilde{\mu}_k = \mu(\mathbf{x}_k, \boldsymbol{\theta})$, and the estimator of $\mathbf{A}(\boldsymbol{\theta}^*)$ is based on $\tilde{\mathbf{I}}(\tilde{\boldsymbol{\theta}})$.

For stratified random sampling and stratified multi-stage sampling, we can also use the jackknife variance estimator $\hat{\mathbf{V}}_J(\tilde{\mathbf{S}}_2)$, similar to (3.3) and (3.8). This requires the computation of $\tilde{\mathbf{S}}_{2(lj)}$ - values which are obtained in the same manner as $\tilde{\mathbf{S}}_2$, using the jackknife weights $\tilde{w}_{hi(lj)}$ for stratified random sampling and $\tilde{w}_{hik(lj)}$ for stratified multi-stage sampling. We replace $\{\hat{\boldsymbol{\theta}}_{(lj)}, \hat{\boldsymbol{\theta}}\}$ by $\{\tilde{\mathbf{S}}_{(lj)}, \tilde{\mathbf{S}}_2\}$ in (3.3) and (3.8) to get $\hat{\mathbf{V}}_J(\tilde{\mathbf{S}_2})$. The jackknife quasi-score test resulting from $\hat{\mathbf{V}}_J(\tilde{\mathbf{S}_2})$ is invariant to reparameterizations, as in the case of tests based on the linearization variance estimator using $\mathbf{A}(\tilde{\boldsymbol{\theta}})$ based on (4.9).

5. Computer Implementation

It is a common practice for statistical agencies to use auxiliary data at the estimation stage 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: LINWEIGHT (Bethlehem and Keller, 1987), PC-CARP (Schnell, Kennedy, Sullivan, Park and Fuller, 1988), SUDAAN (Shah, Lavange, Barnwell, Killinger, Wheeless, 1989), WESVAR (Brick and Morganstein, 1996), CLAN (Anderson and Nordberg, 1994;2000), BASCULA (Nieuwenbreck and Hofman, 2000) and GES (Estevao, Hidiroglou and Särndal, 1995). 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 existing estimation packages. We will assume that the estimation package has the capability to calculate the adjustment factors, $a_k(s)$ (see equation (2.2)), given some auxiliary data $\mathbf{z}_k$ and a variance estimate, $\hat{\mathbf{V}}(\hat{Y}_b) = \mathbf{v}(\mathbf{y}_k)$, of the basic estimate $\hat{\mathbf{Y}}_b$, for a given sample design. This variance estimator will be used as the operator $\mathbf{v}(\cdot)$, as explained in (2.5). With these assumptions, the estimating equations approach may be implemented as follows:

1. Given the design weights, w_k, and the auxiliary data, $\mathbf{z}_k$, the estimation package can be used to compute the adjustment factors $a_k(s)$, given by (2.2), and the resulting calibration weights, $\tilde{w}_k(s)$.

2. An additional computer program is necessary to compute the estimate, $\hat{\boldsymbol{\theta}}$, of the parameter of interest, $\boldsymbol{\theta}_N$, using the estimating equation approach. The parameter of interest is defined by the census estimating equations (2.1). These estimating equations are completely specified by the $\mathbf{u}_k(\boldsymbol{\theta})$-terms. Given the $\mathbf{u}_k(\boldsymbol{\theta})$-terms, the calibration weights, $\tilde{w}_k(s)$, and the necessary data, we can then define the sample estimating equations as $\hat{\mathbf{S}}(\boldsymbol{\theta}) = \sum_{k \in s} \tilde{w}_k(s)\mathbf{u}_k(\boldsymbol{\theta})$. Finally, the program needs to apply the Newton-Raphson (N-R) algorithm to solve the sample estimating equations to obtain $\hat{\boldsymbol{\theta}}$. As a by product to the N-R method, the $\hat{\mathbf{J}}(\boldsymbol{\theta})$ 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 $\mathbf{u}_k(\boldsymbol{\theta})$-terms.

3. Next, given $\hat{\boldsymbol{\theta}}$, the $\mathbf{u}_k(\boldsymbol{\theta})$-terms evaluated at $\hat{\boldsymbol{\theta}}$ and the necessary data, the estimation package can be used again to compute the residuals $\mathbf{e}_k^*(\boldsymbol{\theta})$.

4. Finally, a computer program is needed to calculate the synthetic residuals $a_k(s)\tilde{\mathbf{e}}_k(\hat{\boldsymbol{\theta}})$. The estimation package can then be used to apply the operator $\mathbf{v}(\cdot)$ to these synthetic residuals to get the Taylor linearization variance estimate (or the jackknife linearization variance estimate where applicable) of $\hat{\boldsymbol{\theta}}$. Wald tests can be implemented using $\hat{\boldsymbol{\theta}}$ and the variance estimate of $\hat{\boldsymbol{\theta}}$. Similar steps are needed for implementing quasi-score tests.

References

Anderson, C. and Nordberg, L. (1994). A method for variance estimation of non-linear functions of totals in surveys. Theory and software implementation. *J. Off. Statist.*, **10**, 395-405.

Anderson, C. and Nordberg, L. (2000). CLAN: A SAS-program for computation of point and standard error estimates in sample surveys. *Proc. Second Int. Conf. on Establishment Surveys*, American Statistical Association, 657-666.

Bethlehem K.G. and Keller, W.K. (1987). Linear weighting of sample survey data. *J. Off. Statist.*, **3**, 141-153.

BINDER, D.A. (1983). On the variance of asymptotically normal estimators from complex surveys. *Int. Statist. Rev.*, **51**, 279-292.

BOOS, D.D. (1992). On generalized score tests. *Amer. Statist.*, **46**, 327-333.

BRICK, J.M. and MORGANSTEIN, D.R. (1996). WesVarPC: Software for computing variance estimates from complex designs. *Proc. Bureau of the Census Annual Research Conference*, 861-866.

DEVILLE, J.C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, **25**, 193-203.

DEVILLE, J.C. and SÄRNDAL, C.E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.*, **87**, 376-382.

ESTEVAO, V., HIDIROGLOU, M.A. and SÄRNDAL, C.E. (1995). Methodological principles for a generalized estimation system at Statistics Canada. *J. Off. Statist.*, **11**, 181-204.

GODAMBE, V.P. and THOMPSON, M.E. (1986). Parameters of superpopulation and survey population: their relationship and estimation. *Int. Statist. Rev.*, **54**, 127-138.

HIDIROGLOU, M.A., RAO, J.N.K. and YUNG, W. (1999). Variance computation for complex surveys using estimating equations. *Proc. Survey Meth. Sec.*, Statistical Society of Canada, 3-9.

KORN, E.L. and GRAUBARD, B.I. (1998). Variance estimation for super-population parameters. *Statist. Sinica*, **8**, 131-151.

KOVACEVIC, M.S. and BINDER, D.A. (1997). Variance estimation for measures of income inequality and polarization - the estimating equations approach. *J. Off. Statist.*, **13**, 41-58.

NIEUWENBRICK, R. and HOFMAN, L. (2000). Towards a generalized weighting system. *Proc. Second Int. Conf. on Establishment Surveys*, American Statistical Association, 667-676.

PFEFFERMANN, D. (1993). The role of sampling weights when modeling survey data. *Int. Statist. Rev.*, **61**, 317-337.

RAO, C.R. (1948). Large sample tests of statistical hypothesis concerning several parameters with applications to problems of estimation. *Proc. Cambridge Phil. Soc.*, **44**, 50-57.

RAO, J.N.K., SCOTT A.J. and SKINNER, C.J. (1998). Quasi-score test with survey data. *Statist. Sinica*, **8**, 1059-1070.

ROYALL, R.M. and CUMBERLAND, W.G. (1981). An empirical study of the ratio estimator and estimators of its variance. *J. Amer. Statist. Assoc.*, **76**, 66-77.

RUBIN-BLEUER, S. and SCHIOPU-KRATINA, I. (2001). Joint design and model based inference for fintite populations. *Tech. Rep.*, Business Survey Methods Division, Statistics Canada.

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.

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.

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.

Silva, P.L.N. (1996). *Utilizing Auxiliary Information in Sample Survey Estimation and Analysis*. Unpublished Ph.D. thesis, Department of Social Statistics, Univerity of Southampton.

Skinner, C.J., Holt, D. and Smith, T.M.F. (Eds.) (1989). *Analysis of Complex Surveys*. Wiley, New York.

Valliant, R. (1993). Poststratification and conditional variance estimation. *J. Amer. Statist. Assoc.*, **88**, 89-96.

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

J.N.K. Rao
School of Mathematics and Statistics
Carleton University
1125 Colonel By Drive
Room 4302, Herzberg Building
Ottawa, Ontario K1S 5B6, CANADA
E-mail: jrao@math.carleton.ca

W. Yung
Statistics canada
11-O, R.H. Coats Building
Tunney's Pasture
Ottawa, Ontario K1A 0T6, CANADA
E-mail: yungwes@statcan.ca

M.A. Hidiroglou
Statistics canada
11-A, R.H. Coats Building
Tunney's Pasture
Ottawa, Ontario K1A 0T6, CANADA
E-mail: hidirog@statcan.ca