

VARIANCE ESTIMATION FOR ESTIMATING EQUATIONS IN THE PRESENCE OF MISSING VALUES

Wesley Yung, Michel Hidiroglou and J.N.K. Rao¹

ABSTRACT

Population parameters of interest can be expressed as solutions to appropriate population estimating equations. Sample estimates of these parameters can be obtained by solving suitable sample estimating equations. Hidiroglou, Rao and Yung (1999) showed how variance estimates can be obtained for solutions of estimating equations which incorporates weighting adjustments due to auxiliary information, such as Generalized Regression (GREG) estimation weights. In this paper, we extend the results of Hidiroglou, et al. (2000) to the case where weight adjustments are also made for unit nonresponse or imputation methods are applied to account for missing item values.

KEY WORDS: Estimating equations; Imputation; Jackknife linearization; Variance estimation.

RÉSUMÉ

Des paramètres d'intérêt d'une population peuvent être représentés comme des solutions aux équations d'estimations de la population appropriées. Des estimations d'échantillon de ces paramètres peuvent être obtenues par la solution d'équations d'estimation d'échantillon appropriées. Hidiroglou, Rao et Yung (1999) ont montré comment des estimateurs de la variance peuvent être obtenus pour la solution des équations d'estimation qui incorporent des ajustements pondérés dus à l'information auxiliaire, telle que l'estimation pondérée généralisée de la régression (GREG). Dans cet article, nous étendons les résultats de Hidiroglou et al. (2000) au cas où des ajustements pondérés sont également faits pour la non-réponse ou quand des méthodes d'imputation sont appliquées pour expliquer les valeurs manquantes.

MOTS CLÉS : Équations d'estimations; estimation de la variance; imputation; jackknife linéarisé.

1. INTRODUCTION

Parameters of interest that are estimated from sample survey data are either simple or complex. Simple parameters such as totals, ratios or proportions are usually used for descriptive purposes. On the other hand, complex parameters such as linear or logistic regression vectors and parameters of log linear models are used to obtain a better understanding of relationships that hold within the population of interest. The estimating equations approach provides a unified method for estimating both simple and complex parameters from survey data. Binder (1983) and Godambe and Thompson (1986) showed how linear or non-linear parameters of interest can be expressed as solutions to suitably defined "census" estimating equations. Parameter estimates are then obtained by solving sample estimating equations that

involve design weights, as well as estimation weights based on auxiliary information.

Binder (1983) and Hidiroglou, Rao and Yung (1999) have developed variance estimators for parameter estimates obtained from estimating equations using Taylor linearization and jackknife variance estimation techniques. In addition, Hidiroglou et al. (1999) obtained linearization type variance estimators by linearizing the jackknife variance estimator. The resulting variance estimators incorporate the estimation weights as well as synthetic residuals obtained by regressing the components of the estimating function on the auxiliary variables. Hidiroglou, Rao and Yung (2000) extended their 1999 results to the case of weight adjustments to account for unit nonresponse. In this paper, we extend the results of Hidiroglou et al. (2000) to the case where

¹ Wesley Yung and Michel Hidiroglou, Business Survey Methods Division, 11-O.R.H. Coats Building, Statistics Canada, Ottawa, Ontario, K1A-0T6, yungwes@statcan.ca; J.N.K. Rao, School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, K1S-5B6.

imputation has been used to account for item nonresponse. We derive jackknife and jackknife linearization variance estimators for these parameters under a stratified multistage design.

Section 2 presents the notion of how population estimating equations can be used to define parameters of interest and provides a few examples of how this approach can generate some commonly known parameters in survey sampling. Section 3 shows how sampling and estimation weights can be incorporated into sample estimating equations to produce estimates of the parameters of interest. Variance estimation under full response is given in Section 4, while the nonresponse case is addressed in Section 5. Finally, some summary comments are given in Section 6.

2. CENSUS ESTIMATING EQUATIONS

Suppose that the finite population U is of size N and that for each unit k , we have a P -vector of explanatory variables, \mathbf{x}_k and a response variable y_k . We also assume that for a given \mathbf{x} , the y -variable is generated by a random process with $E_m(y_k) = \mu_k = \mu(\mathbf{x}_k, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a P -vector of parameters and E_m denotes model expectation. Further, denote the “working” variance of y_k by $V_m(y_k) = V_{0k} = \sigma^2 V_0(y_k)$ for $k \in U$, where V_m denotes model variance. A census parameter, $\boldsymbol{\theta}_N$, is defined as the solution to the census estimating equation

$$S(\boldsymbol{\theta}) = \sum_U \mathbf{u}_k(\boldsymbol{\theta}) = \boldsymbol{\theta} \quad (2.1)$$

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

$$u_{kp}(\boldsymbol{\theta}) = \left(\partial \mu_k / \partial \theta_p \right) [(y_k - \mu_k) / V_{0k}] \quad (2.2)$$

for $p=1, \dots, P$.

The estimating equations approach can be used to generate most of the commonly used census parameters, $\boldsymbol{\theta}_N$, such as the mean, ratio of two totals or logistic regression coefficients. For example, the model for producing the mean of a variable y is given by $E_m(y_k) = \mu_k = \theta$, $V_m(y_k) = \sigma^2$ and $Cov_m(y_k, y_l) = 0$, for $k \neq l$. Using (2.2) leads to $u_k(\theta) = y_k - \theta$ and the census estimating functions

$$S(\theta) = \sum_U (y_k - \theta).$$

The solution to $S(\theta) = 0$ is clearly the population mean, $\theta_N = (1/N) \sum_U y_k$. To generate the parameters of a logistic regression, we use the model

$$E_m(y_k) = \mu_k = \frac{\exp(\boldsymbol{\theta}^T \mathbf{x}_k)}{1 + \exp(\boldsymbol{\theta}^T \mathbf{x}_k)}$$

where y_k is a dichotomous random variable taking on the values of 0 or 1. As a working variance, we take the standard binomial form with $V_{0k} = \mu_k(1 - \mu_k)$, so that $\mathbf{u}_k(\boldsymbol{\theta}) = \mathbf{x}_k(y_k - \mathbf{x}_k^T \boldsymbol{\theta})$ and the estimating functions are given by

$$S(\boldsymbol{\theta}) = \sum_U \mathbf{x}_k(y_k - \mathbf{x}_k^T \boldsymbol{\theta}).$$

The census parameter $\boldsymbol{\theta}_N$ is defined implicitly as the solution of $S(\boldsymbol{\theta}) = \boldsymbol{\theta}$.

3. SAMPLE DESIGN

Suppose that we have a stratified multistage design with L strata and N_h clusters or primary sampling units (PSU's) in stratum h . Cluster (hi) contains m_{hi} ultimate units (elements), with the total number of units in stratum h being $M_h = \sum_{i=1}^{N_h} M_{hi}$ and the total number of units being $M = \sum_{h=1}^L M_h$. A sample of n_h (≥ 2) clusters is sampled from each stratum independently across strata and m_{hi} units are sampled from the i -th selected cluster in the h -th stratum. We assume that subsampling within selected clusters is performed to ensure unbiased estimation of cluster totals. We denote the k -th sampled unit in the i -th sampled cluster of the h -th stratum as (hik) ; $k=1, \dots, m_{hi}$, $i=1, \dots, n_h$; $h=1, \dots, L$.

From the specification of the sampling design, we obtain design weights, w_{hik} , associated with the (hik) -th sampled unit and we observe y_{hik} , the variable of interest. At the estimation stage, auxiliary data is commonly used to improve the precision of sample estimators or to benchmark to known population totals. A commonly used estimator that incorporates auxiliary data is the Generalized Regression (GREG) estimator. Suppose that along with the variable of interest, we observe a vector of auxiliary data, \mathbf{z}_{hik} , for each sampled unit and that $\mathbf{Z} = \sum_U \mathbf{z}_k$, the vector of population totals, is known. The GREG estimator of the total, Y , is given as

$$\hat{Y}_G = \sum_s w_{hik} a_{hik} y_{hik} = \sum_s \tilde{w}_{hik} y_{hik},$$

where \sum_s denotes the summation over all sampled units, $\tilde{w}_{hik} = w_{hik} a_{hik}$ is the GREG adjusted final weight and a_{hik} is the GREG estimation weight defined as

$$a_{hik} = \mathbf{1} + \mathbf{z}_{hik}^T \left(\sum_s w_{hik} \mathbf{z}_{hik} \mathbf{z}_{hik}^T \right)^{-1} (\mathbf{Z} - \hat{\mathbf{Z}}) \quad (3.1)$$

with $\hat{\mathbf{Z}} = \sum_s w_{hik} \mathbf{z}_{hik}$, an estimate of the population total vector, \mathbf{Z} . Note that if the final weights, \tilde{w}_{hik} are used to estimate the population totals for \mathbf{z}_{hik} , one will obtain the population vector \mathbf{Z} .

The estimator, $\hat{\boldsymbol{\theta}}$, of the census parameter $\boldsymbol{\theta}_N$ is obtained by solving the GREG (or sample) estimating equations

$$\hat{\mathbf{S}}(\boldsymbol{\theta}) = \sum_s \tilde{w}_{hik} \mathbf{u}_{hik}(\boldsymbol{\theta}) = \mathbf{0}$$

where $\mathbf{u}_{hik}(\boldsymbol{\theta})$ is defined by (2.2).

For some simple parameters, such as the mean and the ratio, an explicit solution to $\hat{\mathbf{S}}(\boldsymbol{\theta}) = \mathbf{0}$ is available. However, for more complex parameters, such as logistic regression coefficients, it maybe necessary to solve the estimating equations iteratively. A commonly used technique is the Newton-Raphson algorithm, the r -th step of which is given by

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

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

4. VARIANCE ESTIMATION UNDER FULL RESPONSE

We next present some results on variance estimation for parameter estimates obtained through the estimating equations approach. We first consider the full response case and give results from Hidiroglou et al. (1999). In section 5, we consider variance estimation in the presence of missing item values.

Variance estimators for parameter estimates obtained from estimating equations in survey sampling have been proposed in the literature. Binder (1983) used the Taylor linearization approach, while Hidiroglou et al. (1999) used both the Taylor and jackknife approaches. They also obtained jackknife linearization versions of the variance estimators. We present here the jackknife approach and also outline the derivation of the jackknife linearization variance estimator.

To calculate a jackknife variance estimator for $\hat{\boldsymbol{\theta}}$, we first define the jackknife weights when the j -th cluster belonging to the g -th stratum has been deleted as

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

The corresponding GREG estimating equations are

$$\hat{\mathbf{S}}_{(gj)}(\boldsymbol{\theta}) = \sum_s \tilde{w}_{hik(gj)} \mathbf{u}_{hik}(\boldsymbol{\theta}) = \mathbf{0}$$

where $\tilde{w}_{hik(gj)} = w_{hik(gj)} a_{hik(gj)}$ are the jackknife adjusted GREG weights and $a_{hik(gj)}$ is obtained from a_{hik} by replacing the design weights, w_{hik} , by the jackknife weights, $w_{hik(gj)}$, in (3.1).

Now to solve $\hat{\mathbf{S}}_{(gj)}(\boldsymbol{\theta}) = \mathbf{0}$, one can use the Newton-Raphson algorithm until convergence or the one-step jackknife as described in Lipsitz, Dear and Zhao (1994). The one-step jackknife simply uses the full sample estimate $\hat{\boldsymbol{\theta}}$ as the starting point and carries out only one step of the Newton-Raphson algorithm. That is

$$\hat{\boldsymbol{\theta}}_{(gj)} = \hat{\boldsymbol{\theta}} + \mathbf{J}_{(gj)}^{-1}(\hat{\boldsymbol{\theta}}) \hat{\mathbf{S}}_{(gj)}(\hat{\boldsymbol{\theta}}) \quad (4.1)$$

where $\mathbf{J}_{(gj)}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{S}}_{(gj)}(\hat{\boldsymbol{\theta}})$, respectively, are obtained from $\mathbf{J}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})$ by replacing the GREG adjusted weights with the jackknife adjusted GREG weights. The jackknife variance estimator of the covariance matrix of $\hat{\boldsymbol{\theta}}$ is then given by

$$v_J(\hat{\boldsymbol{\theta}}) = \sum_g \frac{n_g - 1}{n_g} \sum_j (\hat{\boldsymbol{\theta}}_{(gj)} - \hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{(gj)} - \hat{\boldsymbol{\theta}})^T. \quad (4.2)$$

To obtain a jackknife linearization variance estimator of $\hat{\boldsymbol{\theta}}$, we use $\mathbf{J}_{(gj)}^{-1}(\hat{\boldsymbol{\theta}}) \approx \mathbf{J}^{-1}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ in (4.1) to get an approximation to $\hat{\boldsymbol{\theta}}_{(gj)} - \hat{\boldsymbol{\theta}}$,

$$\hat{\boldsymbol{\theta}}_{(gj)} - \hat{\boldsymbol{\theta}} \approx \mathbf{J}^{-1}(\hat{\boldsymbol{\theta}}) (\hat{\mathbf{S}}_{(gj)}(\hat{\boldsymbol{\theta}}) - \hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})).$$

Now it follows from (4.2) that

$$v_J(\hat{\boldsymbol{\theta}}) \approx \mathbf{J}^{-1}(\hat{\boldsymbol{\theta}}) v_J(\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})) \mathbf{J}^{-1}(\hat{\boldsymbol{\theta}}) \quad (4.3)$$

where $v_J(\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}}))$ is the jackknife estimator of the covariance matrix of $\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})$. However, noting that $\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})$ is simply a GREG estimator, we can approximate $v_J(\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}}))$ by the jackknife linearization variance estimator (see Yung and Rao, 1996) given by

$$\begin{aligned} v_{JL}(\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})) &= \sum_h \frac{n_h - 1}{n_h} \sum_i (e_{hi} - \bar{e}_h)(e_{hi} - \bar{e}_h)^T \\ &= v(e_{hi}) \end{aligned}$$

where $e_{hi} = \sum_k \tilde{w}_{hik} e_{hik}$, the p -th element of e_{hik} is $e_{hikp} = u_{hikp}(\hat{\boldsymbol{\theta}}) - \hat{\mathbf{B}}_p^T(\hat{\boldsymbol{\theta}}) \mathbf{z}_{hik}$, the vector of regression

coefficients, $\hat{\mathbf{B}}_p$ is $\hat{\mathbf{B}}_p(\hat{\theta}) = \hat{\mathbf{A}}^{-1} \sum_s w_{hik} \mathbf{z}_{hik}^T \mathbf{u}_{hikp}(\hat{\theta})$

with $\hat{\mathbf{A}} = \sum_s w_{hik} \mathbf{z}_{hik} \mathbf{z}_{hik}^T$ and the operator notation $v(e_{hi})$ indicates that the variance estimator v_{JL} depends only on the cluster totals e_{hi} . Finally, replacing $v_J(\hat{\mathbf{S}}(\hat{\theta}))$ in (4.3) by $v_{JL}(\hat{\mathbf{S}}(\hat{\theta}))$ gives

$$v_{JL}(\hat{\theta}) = v(\mathbf{J}^{-1}(\hat{\theta}) e_{hi}). \quad (4.4)$$

5. VARIANCE ESTIMATION UNDER NONRESPONSE

The situation where fully completed questionnaires are obtained from all sampled units is highly unlikely to occur as most surveys suffer from both unit and item nonresponse. Unit nonresponse is usually handled by weighting adjustments, whereas item nonresponse is handled by some form of imputation. Treating the imputed values as true values will lead to valid point estimates, under uniform response within imputation classes, but applying standard variance estimation formulae can lead to serious underestimation if the nonresponse rate is appreciable. We first consider the case of unit nonresponse.

5.1 Unit Nonresponse

Unit nonresponse occurs when none of the survey responses are available for a sampled unit. It arises because of situations such as refusals or untraceable units. It is usually compensated for by weighting adjustments where the weights of the respondents are adjusted so that they represent the nonrespondents. Weighting classes are usually defined based on characteristics known for all sample units and weighting adjustments are performed within each class. For simplicity and ease of notation, we assume a single weighting class, but the extension to multiple weighting classes is straightforward.

The nonresponse adjustment factor is given by

$$d = \frac{\sum_s w_{hik}}{\sum_s w_{hik} \delta_{hik}}$$

where δ_{hik} is the response indicator variable defined as $\delta_{hik} = 1$ if the (hik) -th unit responds and 0 otherwise. The GREG estimation weight under nonresponse adjustment is given by

$$d_{hik}^* = 1 + \mathbf{z}_{hik}^T \left(\sum_s d w_{hik} \delta_{hik} \mathbf{z}_{hik} \mathbf{z}_{hik}^T \right)^{-1} (\mathbf{z} - \hat{\mathbf{Z}}^*)$$

where $\hat{\mathbf{Z}}^* = \sum_s d w_{hik} \mathbf{z}_{hik} \delta_{hik}$. The resulting estimating equations are

$$\hat{\mathbf{S}}_{NR}(\theta) = \sum_s d \tilde{w}_{hik}^* \delta_{hik} \mathbf{u}_{hik}(\theta)$$

where $\tilde{w}_{hik}^* = w_{hik} d_{hik}^*$. The estimator $\hat{\theta}$ is obtained by solving $\hat{\mathbf{S}}_{NR}(\theta) = \mathbf{0}$.

To construct a jackknife variance estimator, we define the estimating functions when the (gj) -th cluster has been deleted as

$$\hat{\mathbf{S}}_{NR(gj)}(\theta) = \sum_s d_{(gj)} \tilde{w}_{hik(gj)}^* \delta_{hik} \mathbf{u}_{hik}(\theta)$$

where

$$d_{(gj)} = \sum_s w_{hik(gj)} / \sum_s w_{hik(gj)} \delta_{hik},$$

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

and

$$a_{hik(gj)}^* = 1 + \mathbf{z}_{hik}^T \left(\sum_s d_{(gj)} w_{hik(gj)} \delta_{hik} \mathbf{z}_{hik} \mathbf{z}_{hik}^T \right)^{-1} (\mathbf{z} - \hat{\mathbf{Z}}_{(gj)}^*).$$

The parameter estimate, $\hat{\theta}_{(gj)}$, is obtained by solving $\hat{\mathbf{S}}_{NR(gj)}(\theta) = \mathbf{0}$ and the usual jackknife variance estimator (see (4.2)) is applied. Hidiroglou et al. (2000) presented the following jackknife linearization variance estimator:

$$v_{JL}(\hat{\theta}) = v(\mathbf{J}^{-1}(\hat{\theta})(e_{gj} + \gamma_{gj} \hat{e})), \quad (5.1)$$

where the p -th element of e_{gj} is

$$e_{gjp} = \sum_k d \tilde{w}_{gjk}^* \delta_{gjk} (\mathbf{u}_{gjkp}(\hat{\theta}) - \hat{\mathbf{B}}_p^T \mathbf{z}_{gjk}),$$

$$\hat{\mathbf{B}}_p = \left(\sum_s d w_{hik} \delta_{hik} \mathbf{z}_{hik} \mathbf{z}_{hik}^T \right)^{-1} \sum_s d w_{hik} \delta_{hik} \mathbf{z}_{hik} \mathbf{u}_{hikp}(\hat{\theta}),$$

$$\gamma_{gj} = \frac{1}{\sum_s w_{hik} \delta_{hik}} \sum_k w_{hik} (1 - d \delta_{hik}) \quad \text{and}$$

$$\hat{e} = \sum_s \tilde{w}_{hik}^* \delta_{hik} (\mathbf{u}_{gjk}(\hat{\theta}) - \hat{\mathbf{B}}^T \mathbf{z}_{gjk}).$$

Note that (5.1) will be larger than the variance estimator under complete response (4.4).

5.2 Item Nonresponse

Item nonresponse occurs when a questionnaire is only partially completed and is usually handled through imputation methods such as mean imputation, ratio imputation, regression imputation or weighted hot deck imputation. Once the imputation is performed, it is common practice to treat the imputed values as true values. While this leads to valid point estimates, it can lead to serious underestimation of the true variance. Rao and Shao (1992) proposed an adjusted jackknife variance estimator to correctly estimate the variance when imputation has been performed. Yung and Rao (2000) extended the Rao-Shao adjusted jackknife to GREG estimators that use poststratification

information. In addition, they derived linearization type variance estimators by linearizing the resulting jackknife variance estimators. In subsections 5.2.1 and 5.2.2 we illustrate how the Rao-Shao adjusted jackknife can be used in the estimating equations framework under mean and weighted hot deck imputation. In addition, we present jackknife linearization versions of the resulting variance estimators. As was done in the unit nonresponse case, we restrict ourselves to a single imputation class.

5.2.1 Mean Imputation

For a missing value, mean imputation imputes the weighted mean for the respondents. That is, we impute missing values by $y_{hik}^* = \tilde{y}_r = \tilde{Y}_r / \tilde{N}_r$ where $\tilde{Y}_r = \sum_s \tilde{w}_{hik} \delta_{hik} y_{hik}$ and $\tilde{N}_r = \sum_s \tilde{w}_{hik} \delta_{hik}$. The resulting estimating equations are given by

$$\hat{S}^*(\theta) = \sum_s \tilde{w}_{hik} \delta_{hik} \mathbf{u}_{hik}(\theta) + \sum_s \tilde{w}_{hik} (1 - \delta_{hik}) \mathbf{u}_{hik}^*(\theta) \quad (5.2)$$

where $\mathbf{u}_{hik}^*(\theta)$ is the same as $\mathbf{u}_{hik}(\theta)$ (see (2.2)) except the y_{hik} value have been replaced by the imputed value, y_{hik}^* , if the (hik) -th unit is a nonrespondent. Solving $\hat{S}^*(\theta) = \theta$ we obtain an imputed estimator, $\hat{\theta}^j$ of θ . The estimator $\hat{\theta}^j$ is not asymptotically unbiased under the uniform response mechanism or under the assumed model $E_m(y_{hik}) = \mu_{hik}$. In the latter case, it is asymptotically valid if $\mu_{hik} = \mu$ for all h, i, k . This follows by noting that $E_m(\hat{S}^*(\theta)) = \theta$ if $\mu_{hik} = \mu$.

To obtain a jackknife variance estimator of $\hat{\theta}^j$, we apply the Rao-Shao adjusted jackknife variance estimator which uses adjusted imputed values to compensate for the imputation. In the case of mean imputation, the p -th element of the adjusted imputed value when the (gj) -th cluster has been deleted is

$$u_{hikp(gj)}^*(\theta) = \left(\frac{\partial \mu_{hik}}{\partial \theta_p} \right) (\tilde{y}_{r(gj)} - \mu_{hik}) / V_{0hik} \quad (5.3)$$

where $\tilde{y}_{r(gj)}$ is obtained from \tilde{y}_r by replacing the GREG adjusted weights, \tilde{w}_{hik} , by the jackknife adjusted GREG weights, $\tilde{w}_{hik(gj)}$. The parameter estimate when the (gj) -th cluster has been deleted, $\hat{\theta}_{(gj)}^j$, is the solution of

$$\hat{S}_{(gj)}^*(\theta) = \sum_s \tilde{w}_{hik(gj)} \delta_{hik} \mathbf{u}_{hik}(\theta) + \sum_s \tilde{w}_{hik(gj)} (1 - \delta_{hik}) \mathbf{u}_{hik(gj)}^*(\theta) = \theta,$$

and the jackknife estimator of the covariance matrix is the usual jackknife variance estimator (see (4.2)).

Yung and Rao (2000) obtained jackknife linearization variance estimators for GREG estimators of totals under imputation by linearizing the Rao-Shao adjusted jackknife. Using similar techniques, we obtain a jackknife linearization variance estimator for $\hat{\theta}$ as

$$v_{JL}(\hat{\theta}^j) = v(\mathbf{J}^{-1}(\hat{\theta}^j)[\mathbf{e}_{gj} + \hat{\Gamma}D_{gj}]),$$

where the p -th element of \mathbf{e}_{gj} is

$$e_{gp}(\hat{\theta}^j) = \sum_k \tilde{w}_{gjk} (\tilde{u}_{gjkp}(\hat{\theta}^j) - \hat{\mathbf{B}}_p^T \mathbf{z}_{gjk}),$$

$$D_{gj} = \sum_k \tilde{w}_{gjk} \left[(\delta_{gjk} y_{gjk} - \hat{\mathbf{B}}_3^T \mathbf{z}_{gjk}) - \tilde{Y}_r / \tilde{N}_r (\delta_{gjk} - \hat{\mathbf{B}}_4^T \mathbf{z}_{gjk}) \right]$$

$$\hat{\mathbf{B}}_3 = \hat{\mathbf{A}}^{-1} \sum_s w_{hik} \delta_{hik} y_{hik} \mathbf{z}_{hik},$$

$$\hat{\mathbf{B}}_4 = \hat{\mathbf{A}}^{-1} \sum_s w_{hik} \delta_{hik} \mathbf{z}_{hik},$$

$$\tilde{u}_{gjk}(\hat{\theta}^j) = \begin{cases} \mathbf{u}_{gjk}(\hat{\theta}^j) & \text{if the } (gjk)\text{-th sampled unit responds} \\ \mathbf{u}_{gjk}^*(\hat{\theta}^j) & \text{otherwise} \end{cases}$$

and

$$\hat{\Gamma} = \sum_s \tilde{w}_{hik} (1 - \delta_{hik}) \mathbf{c}_{hik}$$

with the p -th element of \mathbf{c}_{hik} being $c_{hikp} = \partial \mu_{hik} / \partial \theta_p$ evaluated at $\theta = \hat{\theta}^j$.

5.2.2 Hot Deck Imputation

In hot deck imputation, missing values, y_{hik} , are replaced by donor values, y_{hik}^* , where donors are selected with replacement from the respondents with probabilities proportional to their GREG adjusted weights. The resulting estimating equation is the same as (5.2) and the parameter estimate $\hat{\theta}^j$ is obtained in the same manner. The p -th element of the adjusted imputed value for the Rao-Shao adjusted jackknife is

$$u_{hikp(gj)}^*(\theta) = \left(\frac{\partial \mu_{hik}}{\partial \theta_p} \right) (y_{hik}^* + \tilde{y}_{r(gj)} - \tilde{y}_r - \mu_{hik}) / V_{0hik}$$

where $\tilde{y}_{r(gj)}$ and \tilde{y}_r are as previously defined under mean imputation. In the case of mean imputation, the imputed values are identical to \tilde{y}_r so the adjusted imputed value reduces to (5.3). Under hot deck imputation, if we take the expectation with respect to the imputation process we are, on average, imputing the sample mean. That is why the adjusted imputed values under mean imputation and hot deck imputation are similar.

Once the adjusted imputed values are defined, we use the same procedure as under mean imputation to obtain $\hat{\theta}_{(gj)}^j$ and finally the jackknife variance

estimator $v_j(\hat{\theta}^j)$. The jackknife linearization variance estimator is then given by

$$v_{JL}(\hat{\theta}^j) = v\left(J^{-1}(\hat{\theta})[e_{gj} + \hat{\Gamma} D_{gj} + \tilde{e}_{gj}]\right),$$

where e_{gj} , $\hat{\Gamma}$ and D_{gj} are as previously defined and the p -th element of \tilde{e}_{gj} is

$$\tilde{e}_{gjp} = \sum_k \tilde{w}_{gjk}^* (t_{gjkp}^* - \hat{B}_{5p}^T z_{gjk}),$$

$$\hat{B}_{5p} = \hat{A}^{-1} \sum_s w_{hik} \delta_{hik} z_{hik}^T t_{hikp}^*$$

and the p -th element of t_{hik}^* is

$$t_{hikp}^* = (1 - \delta_{hik}) c_{hikp} (y_{hik}^* - \tilde{y}_r) / V_{0hik}.$$

The pseudo-error term in the jackknife linearization variance estimator under hot deck imputation is exactly the same as that obtained under mean imputation except for the extra term \tilde{e}_{gj}^* which is due to the stochastic portion of the hot deck imputation process. This is not unexpected due to the similarities between mean imputation and hot deck imputation mentioned earlier.

6. SUMMARY

We have shown that the estimating equations approach provides a powerful procedure to define and estimate parameters from survey samples. By defining suitable Census estimating equations, sample estimating equations can be obtained leading to parameter estimates. In addition, under certain conditions variance estimation can be handled by the jackknife variance estimation method and can be used in the presence of missing values. In addition to the jackknife method, linearization type variance estimators have been derived under both the full response case and the nonresponse case where unit or item nonresponse has occurred. Work is currently being performed to relax the conditions under which the jackknife variance estimator is valid.

REFERENCES

- Binder, D.A. (1983). On the variance of asymptotically normal estimators from complex surveys. *International Statistical Review*, **51**, 279-292.
- 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., Rao, J.N.K. and Yung, W. (1999). Variance Computation for Complex Surveys using Estimating Equations. *Proceedings of the Survey Methods Section*, Statistical Society of Canada, 3-9.
- Hidiroglou, M., Rao, J.N.K. and Yung, W. (2000). Variance Estimation and Quasi-Score Tests for Complex Surveys using Estimating Equations. *Proceedings of the Survey Research Methods Section*, American Statistical Association, to appear.
- 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.
- Rao, J.N.K. and Shao, J. (1992). Jackknife Variance Estimation with Survey Data under Hot Deck Imputation. *Biometrika*, **79**, 811-822.
- Yung, W. and Rao, J.N.K. (1996). Jackknife Linearization Variance Estimators under Stratified Multistage Sampling. *Survey Methodology*, **22**, 23-31.
- Yung, W. and Rao, J.N.K. (2000). Jackknife Variance Estimation under Imputation for Estimators using Poststratification Information. *Journal of the American Statistical Association*, **95**, 903-915.