SSC Annual Meeting, May 25-28, 2014
Proceeding of the Survey Methods Section

VARIANCE ESTIMATION FROM COMPLEX SURVEY DATA USING LINERAZATION METHOD: IMPACT OF DAVID BINDER

A. Demnati[1] and J.N.K. Rao[2]

ABSTRACT

David Binder's pioneering 1983 paper published in the *International Statistical Review* provided a unified linearization approach to variance estimation from complex survey data. This important paper stimulated much new research on linearization variance estimation in subsequent years. This talk will review some of that work, including the alternative linearization method of Demnati and Rao (2004, 2010). We apply our approach to the estimation of total variance of calibration estimators of the generalized linear models parameters when missing items have been imputed using either deterministic or random imputation.

KEY WORDS: Best linear predictor; Finite population parameters; Imputation; Item nonresponse; Model parameters; Multiple weight adjustments.

RÉSUMÉ

L'article novateur de David Binder publié en 1983 dans le *International Statistical Review* a présenté une approche unifiée de l'estimation de la variance par linéarisation à partir de données d'enquête complexes. Cet article important a fait souffler un vent nouveau sur la recherche sur l'estimation de la variance par linéarisation dans les années qui ont suivi. Nous passerons en revue certains de ces travaux, y compris la méthode de rechange par la linéarisation de Demnati et Rao (2004, 2010). Nous appliquons notre approche à l'estimation de la variance totale des estimateurs par calage pour les paramètres des modèles linéaires généralisées lorsque les réponses manquantes ont été imputées de façon déterministe ou aléatoire.

MOTS CLÉS: Meilleur prédicteur linéaire; paramètres de population finie; imputation; non-réponse partielle; paramètres du modèle; ajustements multiple de poids.

1. INTRODUCTION

Traditionally, statistical agencies collect data to estimate relatively simple finite population quantities such as totals or means. Estimators and associated variance estimators for such parameters are well studied in the literature. Currently, survey data are also used to investigate relationships between variables using statistical models, such as linear and logistic regression models, and interest is in the estimation of the model parameters and associated variance estimators.

Using Taylor series linearization approach, Binder (1983) presented a unified method of estimating the design-based variance of estimators of "census" model parameters based on complex sample designs from finite populations. This method is

[1] A. Demnati, Business Survey Methods Division, Statistics Canada, Ottawa, Canada, Abdellatif.Demnati@statcan.gc.ca

[2] J.N.K. Rao, School of Mathematics and Statistics, Carleton University, Ottawa, Canada, JRao@math.carleton.ca

particularly useful when the parameters are defined implicitly as solutions to estimating equations. Taylor linearization is generally applicable to any sampling design that permits unbiased variance estimation for linear estimators. However, it can lead to multiple variance estimators that are asymptotically design unbiased under repeated sampling. The choice among the variance estimators, therefore, requires other considerations such as (i) approximate unbiasedness for the model variance of the estimator under an assumed model, (ii) validity under a conditional repeated sampling framework. For example, in the context of simple random sampling and the ratio estimator, $\hat{Y}_R = (\bar{y}/\bar{x})X$, of the population total Y, Royall and Cumberland (1981) showed that a commonly used linearization variance estimator, $\vartheta_{LI} = N^2(n^{-1} - N^{-1})s_z^2$, does not track the conditional variance of $\hat{Y}_R$ given $\bar{x}$, unlike the jackknife variance estimator ϑ_J. Here $\bar{y}$ and $\bar{x}$ are the sample means, X is the known population total of the auxiliary variable x, s_z^2 is the sample variance of the residuals $z_k = y_k - (\bar{y}/\bar{x})x_k$ and n denotes the sample size. By linearizing the jackknife variance estimator, ϑ_J, a different linearization variance estimator, $\vartheta_{JL} = (\bar{X}/\bar{x})^2 \vartheta_{LI}$, is obtained. This variance estimator also tracks the conditional variance as well as the unconditional variance, where $\bar{X} = X/N$ is the mean of x. As a result, ϑ_{JL} or ϑ_J may be preferred over ϑ_{LI}. Särndal, Swensson and Wretman (1989) showed that ϑ_{JL} is both asymptotically design unbiased and asymptotically model unbiased in the sense of $E_m(\vartheta_{JL}) \approx Var_m(\hat{Y}_R)$, where $Var_m(\hat{Y}_R)$ is the model variance of $\hat{Y}_R$ under a "ratio model": $E_m(y_k) = \beta x_k$; $k = 1,..., N$ and the y_k's are independent with model variance $Var_m(y_k) = \sigma^2 x_k$, $\sigma^2 > 0$. Thus, ϑ_{JL} is a good choice from either the design-based or the model-based perspective. Binder (1996) presented an elegant "cookbook" approach to Taylor linearization that leads directly to ϑ_{JL}-type linearization variance estimators. He applied the method to smooth functions of estimated totals, $g(\hat{Y}_1,...,\hat{Y}_m)$, generalized regression estimators and the Wilcoxon rank sum statistic.

Demnati and Rao (2004) proposed an alternative Taylor linearization approach to variance estimation that is theoretically justifiable and at the same time leads directly to a ϑ_{JL}-type variance estimator for general designs. They applied the method under the design based approach to a variety of problems, covering regression calibration estimators of a total Y and other estimators defined either explicitly or implicitly as solutions of estimating equations. They obtained a new variance estimator for a general class of calibration estimators that includes generalized raking ratio and generalized regression estimators. They also extended the method to two-phase sampling and obtained a sampling variance estimator that makes fuller use of the first phase sample data compared to traditional linearization variance estimators. Demnati and Rao (2010) studied total variance estimation in the context of finite populations assumed to be generated from superpopulation models and analytical inferences on model parameters are of interest. If the sampling fractions are negligible, then the sampling variance captures almost the entire variation generated by the design and model random processes. However, when the sampling fraction is not negligible, the model variance should be taken into account in order to construct valid inferences on model parameters under both randomization processes.

Demnati and Rao (2002) extended their method to the scalar case of missing responses when weight adjustments for complete nonresponse and imputation for item nonresponse based on smooth functions of observed values, in particular ratio imputation, are used. A standard approach for handling multivariate missing items is to treat one variable at a time and

perform separate imputation. Although separate imputations make the data complete, it does not take advantage of correlations between observed items. Suppose that the respondent is unable to provide a value for each item but rather the respondent is willing to provide the total of two or more items. For example, in business surveys large enterprises are unable to provide value for each combination of industry and geography. However, the total revenue or expense at business level is readily available. In this situation, the separate imputation problem is further complicated by the availability of totals that have to be satisfied by the items. The total value of numerical items must be equal to the sum of its parts. Additionally, if only one item has a missing value then the imputation method should compute the correct value of this variable from the observed values of the other items and their total. This is an example of a situation where the imputation procedure should determine the missing value uniquely from the observed values.

In section 2, we briefly review the Binder method for variance estimation, while in section 3, we give a brief account of the Demnati-Rao (DR) method for total variance estimation. In section 4, we consider estimating equations when calibration and imputation for item nonresponse have been used, and study estimators obtained as solutions to estimating equations. Finally in section 5, we give a preview of our recent work on multivariate imputation for general patterns of missingness under observed control totals. Our multivariate imputation method preserves automatically the observed items while satisfying observed control totals.

2. BINDER'S PIONEERING APPROACH TO VARIANCE ESTIMATION

Consider a finite population of N units identified by a set of indices $P = \{1,...,k,...,N\}$. Usually, many variables (say M variables of interest) are under study for a given survey. Let $\boldsymbol{y}_k = (y_{1k},...,y_{Mk})^T$ be the vector of values of the variable of interest $\boldsymbol{y} = (y_1,...,y_M)^T$ attached to unit k, and let $(\boldsymbol{x}_k^T, \boldsymbol{t}_k^T, \boldsymbol{I}_k^T, \boldsymbol{\chi}_k^T, ...)^T$ be the vector of auxiliary variables attached to unit k, where the superscript T denotes the transpose of a vector or a matrix. We denote a typical variable of interest by y, and a typical auxiliary variable by x. Suppose that the model mean of the response y_k is specified by $E_m(y_k) = \mu_k(\boldsymbol{x}_k^T\boldsymbol{\beta})$, where $\boldsymbol{x}_k = (x_{1k},...,x_{pk})^T$ is a $p \times 1$ vector of explanatory variables, $\boldsymbol{\beta} = (\beta_1,...,\beta_p)^T$ is the $p \times 1$ vector of model parameter and E_m denotes model expectation. Binder (1983) considered the finite population parameter $\boldsymbol{\beta}_N$ defined as solution to "census" estimating equation of the form

$$\boldsymbol{S}_\beta(\boldsymbol{\beta}; \boldsymbol{1}_N, \boldsymbol{y}, \mathbf{A}_x) = \sum_k \boldsymbol{s}_\beta(\boldsymbol{\beta}; y_k, \boldsymbol{x}_k) - \boldsymbol{v}_\beta(\boldsymbol{\beta}) = \mathbf{0}, \tag{2.1}$$

where $\boldsymbol{1}_N$ is the N-vector of 1's, $\mathbf{A}_x$ is a $p \times N$ matrix, with k^{th} column $\boldsymbol{x}_k$, $\boldsymbol{y} = (y_1,...,y_N)^T$, $\sum_k$ denote the sum over all the population units, $\boldsymbol{s}_\beta(\boldsymbol{\beta}; y_k, \boldsymbol{x}_k)$ is a p-dimensional vector-valued function of y_k and $\boldsymbol{x}_k$, and the known function $\boldsymbol{v}_\beta()$ allows for explicitly defined parameters. For linear and logistic regression models, $\boldsymbol{s}_\beta(\boldsymbol{\beta}; y_k, \boldsymbol{x}_k) = \boldsymbol{x}_k(y_k - \mu_k(\boldsymbol{x}_k^T\boldsymbol{\beta}))$, and $\boldsymbol{v}_\beta(\boldsymbol{\beta}) = \mathbf{0}$. For the special case of the finite population total $Y = \sum_k y_k$, $\boldsymbol{s}_\beta(\boldsymbol{\beta}; y_k, \boldsymbol{x}_k) = y_k$, $\boldsymbol{v}_\beta(\boldsymbol{\beta}) = \beta_N$ and $\beta_N = Y$. An estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_N$ is defined by the design weighted estimating equation

$$\boldsymbol{S}_\beta(\boldsymbol{\beta}; \boldsymbol{d}(s), \boldsymbol{y}, \mathbf{A}_x) = \sum_k d_k(s)\boldsymbol{s}_\beta(\boldsymbol{\beta}; y_k, \boldsymbol{x}_k) - \boldsymbol{v}_\beta(\boldsymbol{\beta}) = \mathbf{0}, \tag{2.2}$$

where $d_k(s) = a_k(s)/\pi_k$ denotes the sampling design weights attached to unit k, $a_k(s)$ is the sample membership indicator variable for unit k, $\pi_k = E_p(a_k(s))$ is the inclusion probability for unit k, E_p denotes expectation with respect to the sampling design, and $\boldsymbol{d}(s) = (d_1(s),...,d_N(s))^T$.

Using a Taylor series linearization argument based on the first order approximation, Binder's sampling variance of $\hat{\boldsymbol{\beta}}$ is given by

$$Var_p(\hat{\boldsymbol{\beta}}) = [\boldsymbol{J}(\boldsymbol{\beta}_N)]^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\beta}_N)[\boldsymbol{J}(\boldsymbol{\beta}_N)]^{-1}, \tag{2.3}$$

where $\boldsymbol{J}(\boldsymbol{\beta}) = -\partial \boldsymbol{S}_\beta(\boldsymbol{\beta};\boldsymbol{1}_N,\boldsymbol{y},\mathbf{A}_x)/\partial\boldsymbol{\beta}$, and $\boldsymbol{\Sigma}_U(\boldsymbol{\beta})$ is the sampling variance of $\sum_k d_k(s)\boldsymbol{s}(\boldsymbol{\beta};y_k,\boldsymbol{x}_k)$. Binder's estimator of $Var_p(\hat{\boldsymbol{\beta}})$ is

$$\hat{Var}_p(\hat{\boldsymbol{\beta}}) = [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\beta}})]^{-1}\hat{\boldsymbol{\Sigma}}_U(\hat{\boldsymbol{\beta}})[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\beta}})]^{-1}, \tag{2.4}$$

where $\hat{\boldsymbol{\Sigma}}_U(\hat{\boldsymbol{\beta}})$ is a consistent estimator of $\boldsymbol{\Sigma}_U(\boldsymbol{\beta}_N)$ and $\hat{\boldsymbol{J}}(\boldsymbol{\beta}) = -\partial \boldsymbol{S}_\beta(\boldsymbol{\beta};\boldsymbol{d}(s),\boldsymbol{y},\mathbf{A}_x)/\partial\boldsymbol{\beta}$. Binder (1983) gave regularity conditions for the validity of (2.3) and (2.4), and also showed how the results can be applied to some generalized linear models, including the logistic regression model.

There are a number of competing estimators of the variance that are asymptotically equivalent to (2.4). Binder (1996) presented a "cookbook" approach which produces one of the most favored of those variance estimators. Consider the estimator $\hat{\beta}$ of the finite population parameter β_N which can be expressed as a smooth function

$$\hat{\beta} = g(\hat{Y}_1,...,\hat{Y}_M), \tag{2.5}$$

of estimated totals $\hat{\boldsymbol{Y}} = (\hat{Y}_1,...,\hat{Y}_M)^T$, where $\hat{Y}_i = \sum_k d_k(s)y_{ik}$ is an estimator of the population total $Y_i = \sum_k y_{ik}$, $i = 1,...,M$, and $\beta_N = g(Y_1,...,Y_M)$. Binder' cookbook approach may be summarized as follows:

1) Take the total differential $d\hat{\beta} = \sum_i \frac{dg(\hat{Y}_1,...,\hat{Y}_M)}{d\hat{Y}_i} d\hat{Y}_i$ of $\hat{\beta}$.

2) Replace the total differential $d\hat{\beta}$ by deviations of estimators from their respective population parameters, e.g., $d\hat{\beta}$ is changed to $\hat{\beta} - \beta_N$, and $d\hat{\boldsymbol{Y}}$ is changed to $\hat{\boldsymbol{Y}} - \boldsymbol{Y} = \sum_k d_k(s)\boldsymbol{y}_k - \boldsymbol{Y}$; and so on, to get

$$\hat{\beta} - \beta_N = \sum_i \frac{dg(\hat{Y}_1,...,\hat{Y}_M)}{d\hat{Y}_i}(\sum_k d_k(s)y_{ik} - Y_i). \tag{2.6}$$

3) Rewrite the result of step 2 given by (2.6) as $\hat{\beta} \approx \sum_k z_k d_k(s) + R$, where R does not depends explicitly on the design weight $d_k(s)$.

4) Finally obtain the estimated variance as $\vartheta(\hat{\beta}) \approx \hat{Var}_p(\sum_k d_k(s)z_k) = \hat{Var}_p(\hat{Z})$, using the formula for the variance estimator of the estimated total $\hat{Z}$ under the specified sampling design, where $\hat{Z} = \sum_k d_k(s)z_k$.

The main difference of this formulation from the standard Taylor linearization method is that in the latter approach the partial derivatives are evaluated at their expected values before z_k is derived, and the unknown parameters in the resulting z_k are then replaced by their estimators.

3. DEMNATI-RAO LINERAZITAION METHOD

3.1 The Method

To motivate the DR method, we start with the case of $\hat{\beta}$ defined as a smooth function (2.5) of estimated totals. DR express $\hat{\beta}$ and β_N as $\hat{\beta} = f(\hat{\boldsymbol{d}}(s), \mathbf{A}_y)$ and $\beta_N = f(\boldsymbol{1}_N, \mathbf{A}_y)$, where $\mathbf{A}_y$ is a $M \times N$ matrix, with k^{th} column $\boldsymbol{y}_k = (y_{1k},..., y_{Mk})^T$, $k = 1,..., N$, and $\boldsymbol{1}_N = E_p(\hat{\boldsymbol{d}}(s))$. Note that $\hat{\beta}$ is a function of both $\hat{\boldsymbol{d}}(s)$ and $\mathbf{A}_y$, but we drop $\mathbf{A}_y$ for simplicity and write $\hat{\beta} = f(\hat{\boldsymbol{d}}(s))$. Taylor linearization of $\hat{\beta}$ around $\boldsymbol{Y}$ gives the approximation

$$\hat{\beta} - \beta_N \approx \partial g(\hat{\boldsymbol{Y}}) / \partial \hat{\boldsymbol{Y}}^T |_{\hat{\boldsymbol{Y}}=\boldsymbol{Y}} (\hat{\boldsymbol{Y}} - \boldsymbol{Y}) = \sum_k \tilde{z}_k (d_k(s) - 1), \tag{3.1}$$

where $\tilde{z}_k = f(\boldsymbol{b}) / \partial b_k |_{\boldsymbol{b}=\boldsymbol{1}_N}$ and $\boldsymbol{b}$ is a $N \times 1$ vector of arbitrary real numbers. Note that $(\partial g(\hat{\boldsymbol{Y}}) / \partial \hat{\boldsymbol{Y}}^T)(\hat{\boldsymbol{Y}} - \boldsymbol{Y})$ is equals to the right side of (2.6). The DR sampling variance estimator of $\hat{\beta}$ is given by

$$\vartheta_{DR}(\hat{\beta}) = \vartheta(\sum_k d_k(s) z_k), \tag{3.2}$$

with

$$z_k = \partial f(\boldsymbol{b}) / \partial b_k |_{\boldsymbol{b}=\hat{\boldsymbol{d}}(s)}. \tag{3.3}$$

Demnati and Rao (2004) gave justification of (3.1), (3.2) and (3.3). Note that DR do not first evaluate the partial derivatives of $\partial f(\boldsymbol{b}) / \partial b_k$ at $\boldsymbol{b} = \boldsymbol{1}_N$ and then substitute estimates of the unknown components. The DR method, therefore, is similar in spirit to Binder's cookbook approach.

Estimation of the finite population parameters, $\boldsymbol{\beta}_N = \boldsymbol{h}(\mathbf{A}_y)$, or model parameters, $\boldsymbol{\beta}_M = \boldsymbol{h}(\boldsymbol{\beta})$, under an assumed super-population model on $\boldsymbol{y}$ are often of interest, where $\boldsymbol{\beta}$ is the $p \times 1$ super-population model parameter. We now consider a general formulation of the Demnati and Rao (2004, 2010) approach to deriving Taylor linearization variance estimators. This formulation will cover both finite population (or census) parameters, $\boldsymbol{\beta}_N$, and model parameters, $\boldsymbol{\beta}_M$.

Based on a sample, an estimator, $\hat{\boldsymbol{\beta}}$, is used to estimate both parameters $\boldsymbol{\beta}_N$ and $\boldsymbol{\beta}_M$. Under complete response, the estimator $\hat{\boldsymbol{\beta}}$, obtained as the solution of the estimating equation given by (2.2) is often used as an estimator of the finite population parameter $\boldsymbol{\beta}_N$ defined as the solution to (2.1) and also the model parameter, $\boldsymbol{\beta}_M$, under an assumed model on y.

Let $\hat{\boldsymbol{d}}_k = (\hat{\boldsymbol{d}}_{1k}^T, \hat{\boldsymbol{d}}_{2k}^T,..., \hat{\boldsymbol{d}}_{gk}^T)^T$ be a $G \times 1$ vector of random weights and $\boldsymbol{u}_k$ be a $G \times p$ vector of constants for $k = 1,..., N$. Let $\hat{\boldsymbol{U}} = \sum_k \boldsymbol{u}_k \hat{\boldsymbol{d}}_k$ be a linear estimator and, using an operator notation, let $\vartheta(\boldsymbol{u})$ denote the estimator of variance of $\hat{\boldsymbol{U}}$. We write $\hat{\boldsymbol{\beta}}$ as $f(\mathbf{A}_{\hat{\boldsymbol{d}}})$, where $\mathbf{A}_{\hat{\boldsymbol{d}}}$ is a $G \times N$ matrix with k^{th} column $\hat{\boldsymbol{d}}_k$. The DR linearization variance estimator of $\hat{\boldsymbol{\beta}} = f(\mathbf{A}_{\hat{\boldsymbol{d}}})$ is simply given by $\vartheta_{DR}(\hat{\boldsymbol{\beta}}) = \vartheta(\boldsymbol{z})$, where $\vartheta(\boldsymbol{z})$ is obtained from $\vartheta(\boldsymbol{u})$ by replacing $\boldsymbol{u}_k$ by $\boldsymbol{z}_k = \partial f(\mathbf{A}_{\boldsymbol{b}}) / \partial \boldsymbol{b}_k |_{\mathbf{A}_{\boldsymbol{b}}=\mathbf{A}_{\hat{\boldsymbol{d}}}}$, where $\mathbf{A}_{\boldsymbol{b}}$ is a $G \times N$ matrix of arbitrary real numbers with k^{th} column $\boldsymbol{b}_k = (b_{1k},..., b_{Gk})^T$. The choice of $\mathbf{A}_{\hat{\boldsymbol{d}}}$ depends on the random processes involved. Suppose first that the parameter of interest is $\boldsymbol{\beta}_N$ and we use the estimator given by (2.2). In this case, $g = G = 1$,

$\boldsymbol{d}_k = d_k = d_k(s)$, and $\vartheta(\boldsymbol{u})$ is the estimator of $Var(\sum_k \boldsymbol{u}_k d_k(s))$ with $\boldsymbol{z}_k = [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\beta}})]^{-1}\boldsymbol{s}_\beta(\hat{\boldsymbol{\beta}}; y_k, \boldsymbol{x}_k)$, Suppose on the other hand, we are interested in the model parameter $\boldsymbol{\beta}_M$. In this case, $g = 1$, $G = p$, $\boldsymbol{d}_k = d_k(s)\boldsymbol{s}_\beta(\boldsymbol{\beta}; y_k, \boldsymbol{x}_k)$, and $\vartheta(\boldsymbol{u})$ is the estimator of $Var(\sum_k \boldsymbol{u}_k \boldsymbol{d}_k)$ with $\boldsymbol{z}_k = [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\beta}})]^{-1}$.

The variance estimators associated with the finite population parameter $\boldsymbol{\beta}_N$ and the model parameter $\boldsymbol{\beta}_M$ are different. In the former case, we estimate the variance by an estimator of the design variance $Var_p(\hat{\boldsymbol{\beta}}) = E_p(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_N)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_N)^T$ of $\hat{\boldsymbol{\beta}}$, while in the later case, we estimate the total variance

$$Var(\hat{\boldsymbol{\beta}}) = E_m Var_p(\hat{\boldsymbol{\beta}}) + Var_m E_p(\hat{\boldsymbol{\beta}}),$$

where Var_m denotes the variance with respect to the model. It remains to derive $\boldsymbol{z}_k = \partial f(\mathbf{A}_{\boldsymbol{b}})/\partial \boldsymbol{b}_k \,|_{\mathbf{A}_{\boldsymbol{b}} = \mathbf{A}_{\boldsymbol{d}}}$, when $\hat{\boldsymbol{\beta}}$ is the solution of the estimating equation given by (2.2). One may use a fast approach to derive $\boldsymbol{z}_k$, which consists of two steps: a) first derive $\boldsymbol{z}_{p;k} = \partial f(\boldsymbol{b})/\partial b_k \,|_{\boldsymbol{b}=\boldsymbol{d}(s)}$ where $\hat{\boldsymbol{\beta}} = f(\boldsymbol{d}(s))$; then b) isolate each component $\boldsymbol{z}_{ik}$ of $\boldsymbol{z}_k$ corresponding to component $\boldsymbol{d}_{ik}$ of $\boldsymbol{d}_k$, after approximating $Var(\hat{\boldsymbol{\beta}})$ by $Var(\sum_k \boldsymbol{z}_{p;k} d_k(s))$. Taking the derivative, we get

$$z_{p;k} = \partial f(\boldsymbol{b})/\partial b_k \,|_{\boldsymbol{b}=\boldsymbol{d}(s)} = [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\beta}})]^{-1}\boldsymbol{s}_\beta(\hat{\boldsymbol{\beta}}; y_k, \boldsymbol{x}_k). \tag{3.4}$$

It follows from (3.4) that, when the parameter of interest is the finite population parameter $\boldsymbol{\beta}_N$, the linearized variable is given by (3.4), while in the case where the parameter of interest is the model parameter $\boldsymbol{\beta}_M$, the linearized variable is given $\boldsymbol{z}_k = [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\beta}})]^{-1}$.

3.2 Calibration Estimator

Calibration estimator, $\tilde{\boldsymbol{\beta}}$, is the solution to weighted estimating equations of the form

$$\boldsymbol{S}_\beta(\boldsymbol{\beta}; \boldsymbol{w}(s), \boldsymbol{y}, \mathbf{A}_x) = \mathbf{0},$$

with weights $w_k(s) = d_k(s)F(\boldsymbol{t}_k^T\hat{\boldsymbol{\gamma}})$,

and satisfying the calibration constrains

$$\boldsymbol{S}_\gamma(\boldsymbol{\gamma}; \boldsymbol{d}(s), \boldsymbol{t}) = \sum_k d_k(s)\boldsymbol{s}_\gamma(\boldsymbol{\gamma}; t_k) - \boldsymbol{v}_\gamma(\boldsymbol{\gamma}) = \mathbf{0},$$

with $\boldsymbol{s}_\gamma(\boldsymbol{\gamma}; t_k) = F(\boldsymbol{t}_k^T\hat{\boldsymbol{\gamma}})\boldsymbol{t}_k$ and $\boldsymbol{v}_\gamma(\boldsymbol{\gamma}) = \boldsymbol{T}$, where $\boldsymbol{w}(s) = (w_1(s), ..., w_N(s))^T$, $\boldsymbol{t}_k = (t_{1k}, ..., t_{qk})^T$ and $\boldsymbol{T} = (T_1, ..., T_q)^T$ is the vector of known totals of auxiliary variable $\boldsymbol{t} = (t_1, ..., t_q)^T$. Calibration estimators are widely used in practice. For example, the choice $F(a) = 1 + a$ gives the generalized regression (GREG) weights and $F(a) = \exp(a)$ leads to raking ratio weights. Taking the derivative $\partial f(\boldsymbol{b})/\partial b_k \,|_{\boldsymbol{b}=\boldsymbol{d}(s)}$ with $\tilde{\boldsymbol{\beta}} = f(\boldsymbol{d}(s))$, we get

$$\boldsymbol{z}_{p;k} = [\hat{\boldsymbol{J}}(\tilde{\boldsymbol{\beta}})]^{-1}F(\boldsymbol{t}_k^T\hat{\boldsymbol{\gamma}})\{\boldsymbol{s}_\beta(\tilde{\boldsymbol{\beta}}; y_k, \boldsymbol{x}_k) - \hat{\mathbf{B}}^T(\boldsymbol{s})\boldsymbol{t}_k\}, \tag{3.5}$$

where $\hat{\mathbf{B}}(\boldsymbol{s}) = [\sum_k d_k(s) f(\boldsymbol{t}_k^T\hat{\boldsymbol{\gamma}})\boldsymbol{t}_k\boldsymbol{t}_k^T]^{-1}\sum_k d_k(s) f(\boldsymbol{t}_k^T\hat{\boldsymbol{\gamma}})\boldsymbol{t}_k\boldsymbol{s}_\beta^T(\tilde{\boldsymbol{\beta}}; y_k, \boldsymbol{x}_k)$,

and $f(a) = \partial F(a)/\partial a$.

When the parameter of interest is the finite population parameter $\boldsymbol{\beta}_N$, $g = G = 1$, $\boldsymbol{d}_k = d_k(s)$, and the linearized variable is given by (3.5), while in the case where the parameter of interest is the model parameter $\boldsymbol{\beta}_M$, $g = 2$ $G = 1 + p$, $d_{1k} = d_k(s)$, $\boldsymbol{d}_{2k} = d_k(s)\boldsymbol{s}_\beta(\boldsymbol{\beta}; y_k, \boldsymbol{x}_k)$, and the components of the linearized variable $\boldsymbol{z}_k = (\boldsymbol{z}_{1k}^T, \boldsymbol{z}_{2k}^T)^T$ are given by $\boldsymbol{z}_{1k} = -[\hat{\boldsymbol{J}}(\tilde{\boldsymbol{\beta}})]^{-1} F(\boldsymbol{t}_k^T \hat{\boldsymbol{\gamma}}) \hat{\mathbf{B}}^T(\boldsymbol{s}) \boldsymbol{t}_k$ and $\boldsymbol{z}_{2k} = [\hat{\boldsymbol{J}}(\tilde{\boldsymbol{\beta}})]^{-1} F(\boldsymbol{t}_k^T \hat{\boldsymbol{\gamma}})$.

4. DEMNATI-RAO LINEARIZATION METHOD UNDER IMPUITATION

Item nonresponse occurs in a survey when a sampled unit fails to provide responses to one or more of the survey items. A standard approach for handling missing items is to impute (i.e., fill in) an estimated value for each missing item using an imputation model in combination with an estimate, $\tilde{\boldsymbol{\beta}}$, of the model parameter $\boldsymbol{\beta}$, calculated from the incomplete data. We introduce $\boldsymbol{y}_{\{k\}}^{(o)}$ as the set of observed items of the variables of interest for unit k, and $\boldsymbol{I}_{\{k\}}^{(o)}$ as the set of all observed values attached to unit k related to both the variables of interest and to the auxiliary variables. In case of complete response from unit k $\boldsymbol{y}_{\{k\}}^{(o)} = \{y_{1k}, ..., y_{Mk}\}$, while under complete nonresponse $\boldsymbol{y}_{\{k\}}^{(o)} = \varnothing$, where $\varnothing$ denotes the empty set. After imputation, it is common practice to treat the imputed values as if they were observed and then compute estimator $\tilde{\theta}$ of the parameter of interest θ as in the complete response case, where θ denotes either a finite population parameter θ_N or a model parameter θ_M. Then it remains to assess the accuracy of the estimator $\tilde{\theta}$ in estimating the parameter θ.

4.1 Conditional Mean Imputation Based Estimator

Let $\tilde{y}_k^{(*)} = E_m(y_k \mid \boldsymbol{I}_{\{k\}}^{(o)}; \tilde{\boldsymbol{\beta}})$ be the conditional mean imputed value under the imputation model. Note that $\tilde{y}_k^{(*)} = y_k$ when the item y_k is observed. The previous chapter covers the problem of estimating the variance of $\tilde{\boldsymbol{\beta}}$ solution of the weighted estimating function $\boldsymbol{S}_\beta(\boldsymbol{\beta}; \boldsymbol{w}(s), \boldsymbol{y}, \mathbf{A}_x) = \mathbf{0}$ having zero mean for the k^{th} component at the true model parameter $\boldsymbol{\beta}$, i.e., $E_m\{\boldsymbol{s}_\beta(\boldsymbol{\beta}; y_k, \boldsymbol{x}_k)\} = \mathbf{0}$. In this section estimating equations are used in the more general concept than estimating function which includes weighted log-likelihood estimating function as well as weighted least square estimating functions. A simple example of estimating equation is when one is interested in the overall mean, i.e., $\boldsymbol{S}_\theta(\boldsymbol{\theta}; \boldsymbol{w}(s), \tilde{\boldsymbol{y}}^{(*)}, \mathbf{A}_\chi) = \mathbf{0}$ with $\boldsymbol{s}_\theta(\boldsymbol{\theta}; \tilde{y}_k^{(*)}, \boldsymbol{\chi}_k) = \tilde{y}_k^{(*)} - \theta$ and $\boldsymbol{v}_\theta(\boldsymbol{\theta}) = \mathbf{0}$, when $E_m(y_k) = \mu_k(\boldsymbol{x}_k^T \boldsymbol{\beta})$, where $\boldsymbol{\chi}_k$ is the vector of auxiliary variables associated with $\tilde{y}_k^{(*)}$ and θ. The parameter of interest induced by the estimator $\tilde{\theta}$ is either the finite population parameter solution to $E_r E_p\{\boldsymbol{S}_\theta(\boldsymbol{\theta}; \boldsymbol{w}(s), \tilde{\boldsymbol{y}}^{(*)}, \mathbf{A}_\chi) = \mathbf{0}\}$ or the model parameter solution to $E_m E_r E_p\{\boldsymbol{S}_\theta(\boldsymbol{\theta}; \boldsymbol{w}(s), \tilde{\boldsymbol{y}}^{(*)}, \mathbf{A}_\chi) = \mathbf{0}\}$, where E_r denotes expectation with respect to the response mechanism. Given that interest is in estimating the variance of $\tilde{\theta}$, the vector parameter $\boldsymbol{\beta}$ used for imputation can be seen as a nuisance having $E\{\boldsymbol{s}_\beta(\boldsymbol{\beta}; y_k, \boldsymbol{x}_k)\} = \mathbf{0}$, where $\tilde{\boldsymbol{\beta}}$ is the incomplete-data based estimator solution to

$$\boldsymbol{S}_\beta(\boldsymbol{\beta}; \hat{\boldsymbol{w}}(s), \boldsymbol{y}, \mathbf{A}_x) = \mathbf{0},$$

with

$$\hat{w}_k(s) = w_k(s)(\delta_k / \tilde{\xi}_k),$$

and $\tilde{\boldsymbol{\alpha}}$ in $\tilde{\xi}_k = \xi_k(\tilde{\boldsymbol{\alpha}})$ is the solution to the logistic regression estimating equation

$$\boldsymbol{S}_{\alpha}(\boldsymbol{\alpha};\boldsymbol{w}(s),\boldsymbol{\delta},\mathbf{A}_{I})=\mathbf{0}\,,$$

where $\boldsymbol{\delta}=(\delta_1,...,\delta_N)^T$, δ_k is the response indicator for y_k, i.e., $\delta_k=1$ if y_k is observed and $\delta_k=0$ if y_k is missing, $\boldsymbol{s}_{\alpha}(\boldsymbol{\alpha};\delta_k,\boldsymbol{I}_k)=\boldsymbol{I}_k(\delta_k-\xi_k)$, $\boldsymbol{v}_{\alpha}(\boldsymbol{\alpha})=\mathbf{0}$, $\xi_k=\Pr(\delta_k=1)=logit(\boldsymbol{I}_k^T\boldsymbol{\alpha})$, and $\boldsymbol{I}_k$ is the vector of explanatory variables,

After imputation, the estimator $\tilde{\boldsymbol{\theta}}$ is obtained as the solution to

$$\boldsymbol{S}_{\theta}(\boldsymbol{\theta};\boldsymbol{w}(s),\tilde{\boldsymbol{y}}^{(*)},\mathbf{A}_{\chi})=\mathbf{0}\,.$$

The imputed value $\tilde{y}_k^{(*)}=E_m(y_k\,|\,\boldsymbol{I}_{\{k\}}^{(o)};\tilde{\boldsymbol{\beta}})$ can be rewritten as

$$\tilde{y}_k^{(*)}=\delta_k y_k+(1-\delta_k)\tilde{y}_k^{(p)}\,,$$

with the missing values replaced by predicted values given by

$$\tilde{y}_k^{(p)}=E_m(y_k\,|\,\boldsymbol{I}_{\{k\}}^{(o)}\setminus\{y_k\};\tilde{\boldsymbol{\beta}})\,,$$

available for all sampled units. Note that $\tilde{y}_k^{(*)}=\tilde{y}_k^{(p)}$ for non respondents and $\tilde{y}_k^{(*)}=y_k$ for respondents.

4.2 Random Imputation Based Estimator

Suppose we can sample random values $\tilde{y}_k^{(R)}$ from the conditional distribution with mean $\tilde{y}_k^{(*)}$ and variance $Var_m(y_k\,|\,\boldsymbol{I}_{\{k\}}^{(o)},\tilde{\boldsymbol{\beta}})$, so that $\tilde{y}_k^{(R)}$ are independent observations from the known conditional distribution $f(\tilde{y}_k^{(*)},Var_m(y_k\,|\,\boldsymbol{I}_{\{k\}}^{(o)},\tilde{\boldsymbol{\beta}}))$. Under random imputation, the estimator $\tilde{\boldsymbol{\theta}}^{(R)}$ is obtained as the solution to

$$\boldsymbol{S}_{\theta}(\boldsymbol{\theta};\boldsymbol{w}(s),\tilde{\boldsymbol{y}}^{(R)},\mathbf{A}_{\chi})=\mathbf{0}\,.$$

We may write $\tilde{y}_k^{(R)}$ as $\tilde{y}_k^{(R)}=\tilde{y}_k^{(*)}+\varepsilon_k$, where $\varepsilon_k=\tilde{y}_k^{(R)}-\tilde{y}_k^{(*)}$ with $E_R(\varepsilon_k)=0$, and E_R denotes expectation with respect to the random draw for imputation. In case of conditional mean imputation $\varepsilon_k=0$.

5. PREVIEW ON MULTIVARIATE CONDITIONAL IMPUTATION

We now give a preview of the recent work of Demnati and Rao (2014) on variance estimation under multivariate conditional imputation for general patterns of missingness under observed control totals. Our multivariate imputation method preserves automatically the observed items while satisfying observed control totals.

5.1 Missing Multivariate Items Situation

Table 1 presents a simple example of a missing data situation for the first six units of a population of size $N=500$. Three variables and their total, labeled y_1, y_2, y_3 and t, are observed, but some of the y_1, y_2, y_3 and the t values marked blank in shaded cells are missing. The complete population is generated from a 3-variate Normal distribution with mean $\boldsymbol{\mu}=(5,10,15)^T$, variance $\boldsymbol{\sigma}^2=(5,10,15)^T$ and coefficient of correlation $\boldsymbol{\rho}=(\rho_{12},\rho_{13},\rho_{23})^T=(.5,.7,.9)^T$. The probabilities of response are set to .5 for

the four variables. The first unit of Table 1 provided complete response, while no information was received from unit 6. Note that the missing item from unit 2 can be determined exactly from the reported values.

Table 1: Observed values

Unit	y_1	y_2	y_3	t
1	7.12	19.71	34.13	60.96
2		-13.53	-7.74	-19.01
3	5.83			30.77
4				19.50
5	10.04			
6				
⋮				

5.2. Observed Data

Consider the case where the data vector of interest $\boldsymbol{y}_k = (y_{1k},...,y_{Mk})^T$ for unit k is not observed but instead we observe a $m_k \times 1$ vector $\boldsymbol{I}_k^{(o)}$ which is a linear combination of $\boldsymbol{y}_k$,

$$\boldsymbol{I}_k^{(o)} = \boldsymbol{L}_k^T \boldsymbol{y}_k ,$$

where $\boldsymbol{L}_k$ is a $M \times m_k$ full rank column matrix. For example if we observed the subset (y_{1k}, y_{3k}) only, then the $2 \times M$ matrix $\boldsymbol{L}_k^T$ is given by

$$\boldsymbol{L}_k^T = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \end{bmatrix}.$$

In some situations we observe the sum of some components of $\boldsymbol{y}_k$. For example, the sum of all components is obtained by using the $1 \times M$ vector $\boldsymbol{L}_k^T$ given by $\boldsymbol{L}_k^T = [1 \ \dots \ 1] \equiv \boldsymbol{1}_M^T$, where $\boldsymbol{1}_M$ is the $M \times 1$ vector of 1's.

Note that $\boldsymbol{I}_k^{(o)}$ varies from one unit to another. The intention of the imputation process is to provide either a complete vector $\hat{\boldsymbol{y}}_k^{(*)} = (\hat{y}_{1k}^{(*)},...,\hat{y}_{Mk}^{(*)})^T$ through conditional mean imputation or a complete vector $\hat{\boldsymbol{y}}_k^{(R)} = (\hat{y}_{1k}^{(R)},...,\hat{y}_{Mk}^{(R)})^T$ through random imputation as estimate for $\boldsymbol{y}_k$ given the observed information $\boldsymbol{I}_k^{(o)}$, with $\hat{y}_{ik}^{(*)} = \hat{y}_{ik}^{(R)} = y_{ik}$ if item i for unit k is observed explicitly or implicitly. Our imputed value of $\boldsymbol{y}_k$ given $\boldsymbol{I}_k^{(o)}$ will of course be some function of $\boldsymbol{I}_k^{(o)}$, say $\boldsymbol{y}(\boldsymbol{I}_k^{(o)})$.

5.3. Conditional Mean Imputation for Item Nonresponse

Suppose that $(\boldsymbol{y}_k, \boldsymbol{I}_k^{(o)})$ are jointly distributed variables and we wish to predict the $M \times 1$ vector $\boldsymbol{y}_k$ from the $m_k \times 1$ observed vector $\boldsymbol{I}_k^{(o)}$. One can in fact find the best estimator $\boldsymbol{y}^{(*)}(\boldsymbol{I}_k^{(o)})$ in the sense of minimizing MSE over all estimators. Except for the case of multivariate normal distributions, the conditional expectation $\boldsymbol{y}^{(*)}(\boldsymbol{I}_k^{(o)})$ could be a complicated nonlinear function of $\boldsymbol{I}_k^{(o)}$. So we restrict the class of estimators to the so-called linear estimators,

$$\boldsymbol{y}_L^{(*)}(\boldsymbol{l}_k^{(o)}) = \boldsymbol{\mu}_k + \boldsymbol{Q}_{\boldsymbol{y}_k \boldsymbol{l}_k^{(o)}} \boldsymbol{Q}_{\boldsymbol{l}_k^{(o)} \boldsymbol{l}_k^{(o)}}^{-1} (\boldsymbol{l}_k^{(o)} - \boldsymbol{\mu}_{\boldsymbol{l}^{(o)};k}) \equiv E_{mL}(\boldsymbol{y}_k \mid \boldsymbol{l}_k^{(o)}), \tag{5.1}$$

with $$MSE[\boldsymbol{y}_L^{(*)}(\boldsymbol{l}_k^{(o)})] = Var_m(\boldsymbol{y}_k) - \boldsymbol{Q}_{\boldsymbol{y}_k \boldsymbol{l}_k^{(o)}} \boldsymbol{Q}_{\boldsymbol{l}_k^{(o)} \boldsymbol{l}_k^{(o)}}^{-1} \boldsymbol{Q}_{\boldsymbol{l}_k^{(o)} \boldsymbol{y}_k} \equiv Var_{mL}(\boldsymbol{y}_k \mid \boldsymbol{l}_k^{(o)}),$$

where $\boldsymbol{\mu}_k = E_m(\boldsymbol{y}_k)$, $\boldsymbol{\mu}_{\boldsymbol{l}^{(o)};k} = E_m(\boldsymbol{l}_k^{(o)})$, $\boldsymbol{Q}_{\boldsymbol{y}_k \boldsymbol{l}_k^{(o)}} = Cov_m(\boldsymbol{y}_k, \boldsymbol{l}_k^{(o)})$ and $\boldsymbol{Q}_{\boldsymbol{l}_k^{(o)} \boldsymbol{l}_k^{(o)}} = Cov_m(\boldsymbol{l}_k^{(o)}, \boldsymbol{l}_k^{(o)})$. As shown by Goldberger (1962), the linear estimator given by (5.1) is the best linear unbiased predictor of $\boldsymbol{y}_k$ under the general linear model. The relative error, $\boldsymbol{re}_k$, of $\boldsymbol{y}_k$ based on the knowledge $\boldsymbol{l}_k^{(o)}$ can be defined by the diagonal elements of the matrix

$$\mathbf{E}_k = [diag Var_m(\boldsymbol{y}_k)]^{-1} diag Var_{mL}(\boldsymbol{y}_k \mid \boldsymbol{l}_k^{(o)}).$$

If $\boldsymbol{l}_{\{k\}}^{(o)} = \varnothing$ then $E_{mL}(\boldsymbol{y}_k \mid \boldsymbol{l}_k^{(o)}) = E_m(\boldsymbol{y}_k)$, $Var_{mL}(\boldsymbol{y}_k \mid \boldsymbol{l}_k^{(o)}) = Var_m(\boldsymbol{y}_k)$ and $\mathbf{E}_k = \mathbf{I}_M$, where $\mathbf{I}_M$ is the $M \times M$ identity matrix. On the other hand, if $\boldsymbol{l}_k^{(o)} = \boldsymbol{y}_k$ then $E_{mL}(\boldsymbol{y}_k \mid \boldsymbol{l}_k^{(o)}) = \boldsymbol{y}_k$, $Var_{mL}(\boldsymbol{y}_k \mid \boldsymbol{l}_k^{(o)}) = \mathbf{0}$ and $\mathbf{E}_k = \mathbf{0}_M$, where $\mathbf{0}_M$ is the $M \times M$ matrix of zero's.

Consider the case of $M = 2$, where $\boldsymbol{y}_k = (y_{1k}, y_{2k})^T$, $k = 1,..., N$, are independent observations generated from a bivariate normal distribution with means $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$, variance $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2)^T$ and coefficient of correlation $\rho_{y_1 y_2}$. We have $\sigma_t^2 = \sigma_1^2 + \sigma_2^2 + 2\rho_{y_1 y_2}\sigma_1\sigma_2$ for $t = y_1 + y_2$, $Cov_m(y_i, t) = \sigma_i^2 + \rho_{y_1 y_2}\sigma_1\sigma_2$ and $\rho_{y_i t}^2 = [\sigma_i + \rho_{y_1 y_2}\sigma_j]^2 / [\sigma_1^2 + \sigma_2^2 + 2\rho_{y_1 y_2}\sigma_1\sigma_2]$ for $i \neq j$. Let $\alpha_{i|j} = Cov_m(y_i, y_j) / Var_m(y_j) = \rho_{y_1 y_2}\sigma_i / \sigma_j$ for $i \neq j$, and $\alpha_{i|t} = Cov_m(y_i, t) / \sigma_t^2$. Table 2 reports the imputed values and the relative errors given some patterns of missingness.

Table 2: Imputed value, $y_k^{(*)}$, and relative error given some response patterns when $\boldsymbol{y}_k \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Pattern of Missing Data			$E_m(\boldsymbol{y}_k \mid \boldsymbol{l}_k^{(o)}, \boldsymbol{\beta})$		$\boldsymbol{re}(\boldsymbol{y}_k)$	
y_1	y_2	t	$y_1^{(*)}$	$y_2^{(*)}$	$re(y_1)$	$re(y_2)$
1	1	0	y_1	y_2	0	0
1	0	1	y_1	y_2	0	0
1	0	0	y_1	$\mu_2 + \alpha_{2\|1}(y_{1k} - \mu_1)$	0	$1 - \rho_{y_1 y_2}^2$
0	0	1	$\mu_1 + \alpha_{1\|t}(t_k - \mu_t)$	$\mu_2 + \alpha_{2\|t}(t_k - \mu_t)$	$1 - \rho_{y_1 t}^2$	$1 - \rho_{y_2 t}^2$
0	0	0	μ_1	μ_2	1	1

The left panel of Table 3 shows all the true values related to Table 1. The middle panel shows the imputed values using conditional mean imputation, and the right panel shows the relative error in percentage due to imputation. The middle and right panels are obtained using our method studied in Demnati and Rao (2014), assuming the model parameter is known. Although the responses are generated independently with the same probability of response, it is seen from the right panel of Table 3 that correlations between variables makes approximate rather than fixed and exact response indicators compared to binary variable where the indicators take only zero or one. The unobserved item have a value that ranges in error between zero and one, and this value depends on the observed components and their relationship to the missing item. Although the probability of response is maintained the same for all variables, variable 3 has the lowest mean relative error.

Table 3 : True-, Imputed-values and Relative errors

		True values			Imputed values			Relative errors (%)		
Unit	t	y_1	y_2	y_3	$y_1^{(*)}$	$y_2^{(*)}$	$y_3^{(*)}$	$100re_1$	$100re_2$	$100re_3$
1	Y	7.12	19.71	34.13	7.12	19.71	34.13	0	0	0
2	Y	2.26	-13.53	-7.74	2.26	-13.53	-7.74	0	0	0
3	Y	5.83	12.93	20.01	5.83	13.21	19.73	0	5	2
4	Y	5.98	6.59	6.93	3.61	6.48	9.41	46	13	2
5	N	10.04	14.03	23.22	10.04	15.04	25.58	0	75	51
6	N	-.08	-3.97	-7.23	5	10	15	100	100	100
⋮										
Mean		3.54	10.87	12.89	4.14	11.36	15.80	20.6	15.5	13.35

Demnati and Rao (2014) developed estimators and variance estimators in the case of conditional multivariate imputation for item nonresponse.

REFERENCES

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

Binder, D. A. (1996). "Linearization methods for single phase and two-phase samples: a cookbook approach". *Survey Methodology*, **22**, 17-22.

Demnati, A. and Rao, J. N. K. (2002). "Linearization variance estimators for survey data with missing responses". *Proceeding of the Section Survey Research Methods*, American Statistical Association, 736-740.

Demnati, A. and Rao, J.N.K. (2004). "Linearization variance estimators for survey data (with discussion)". *Survey Methodology*, **30**, 17-34.

Demnati, A. and Rao, J.N.K. (2010). "Linearization variance estimators for model parameters from complex survey data". *Survey Methodology*, **36**, 193-199.

Demnati, A. and Rao, J.N.K. (2014). "Multivariate conditional mean imputation for general pattern of missingness under control totals". In progress.

Goldberger, A.S. (1962). "Best Linear unbiased prediction in the generalized linear regression model". *Journal of the American Statistical Association*, **57**, 369-375.

Royall, R. M. and Cumberland, W. G. (1981). “An empirical study of the ratio estimator and estimators of its variance”. *Journal of the American Statistical Association,* **76**, 66-77.

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.