9-16-2009

A unified approach to linearization variance estimation from survey data after imputation for item nonresponse

Jae Kwang Kim
Iowa State University, jkim@iastate.edu

J. N. K. Rao
Carleton University

Biometrika Advance Access published September 16, 2009

Biometrika (2009), *pp*. 1–16
doi: 10.1093/biomet/asp041

Printed in Great Britain

A unified approach to linearization variance estimation from survey data after imputation for item nonresponse

BY JAE KWANG KIM

Department of Statistics, Iowa State University, Ames, Iowa 50011, U.S.A.
jkim@iastate.edu

AND J. N. K. RAO

School of Mathematics and Statistics, Carleton University, Ottawa, Ontario K1S 5B6, Canada
jrao@math.carleton.ca

SUMMARY

Variance estimation after imputation is an important practical problem in survey sampling. When deterministic imputation or stochastic imputation is used, we show that the variance of the imputed estimator can be consistently estimated by a unifying linearize and reverse approach. We provide some applications of the approach to regression imputation, fractional categorical imputation, multiple imputation and composite imputation. Results from a simulation study, under a factorial structure for the sampling, response and imputation mechanisms, show that the proposed linearization variance estimator performs well in terms of relative bias, assuming a missing at random response mechanism.

Some key words: Composite imputation; Fractional imputation; Imputed estimator; Multiple imputation; Regression imputation.

1. INTRODUCTION

Imputation is a process of assigning values for a missing item y, using observed auxiliary variables $x = (x_1, \ldots, x_p)^{\mathrm{T}}$, to produce a complete dataset. Reasons for conducting imputation are to facilitate analyses using complete data analysis methods, to ensure that the results obtained by different analyses are consistent with one another, and to reduce nonresponse bias. Haziza (2009) provides a comprehensive overview of the imputation methods commonly used in survey sampling.

Variance estimation after imputation is an important practical problem in survey sampling. Treating the imputed values as if observed and then applying the standard variance estimation formula often leads to underestimation. Approaches to variance estimation that account for imputation include the multiple imputation of Rubin (1987), the adjusted jackknife method of Rao & Shao (1992), the population-model approach of Särndal (1992) and Deville & Särndal (1994) and the fractional imputation method of Fuller & Kim (2005).

In this paper, we discuss a unified approach to linearization variance estimation with imputed data under an assumed population model for the item y given x and missing at random mechanism. The parameter in the imputation model is a nuisance parameter in the sense that the main parameter of interest is the population total of y, not the parameters of the imputation model. A regression coefficient in the regression imputation scheme is an example of a nuisance parameter. As the imputed estimator can be viewed as a function of estimated nuisance parameters in the imputation model, Taylor expansion methods can be applied to account for the sampling variability of the estimated nuisance parameter. We call this the linearize and reverse approach because it uses the Taylor linearization of the imputed estimator with respect to the estimated nuisance parameters and then applies the reverse approach that is based on an extended definition of the respondents (Fay, 1992). The implementation of this approach is similar to that of Shao & Steel (1999) and thus can be viewed as an extension of their method. Furthermore, this approach can be easily applied to other problems including domain estimation, multiple imputation and composite imputation involving two or more different imputation methods, depending on the availability of the components of x.

2. DETERMINISTIC IMPUTATION

Consider a finite population of N elements indexed by $U = \{1, \ldots, N\}$. Associated with each element i in the population are two study variables, $x_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ and y_i, where x_i is always observed and y_i is subject to nonresponse. Let s denote the set of indices for the elements in a sample selected by a probability sampling. Under complete response, we consider a design-unbiased estimator for the finite population total $\theta_N = \sum_{i=1}^{N} y_i$ of the form $\hat{\theta}_n = \sum_{i \in s} w_i y_i$, where $w_i = 1/\pi_i$ is the design weight assigned to element i and $\pi_i > 0$ is the inclusion probability for element $i \in U$. We also assume that

$$\hat{V}_n = \sum_{i \in s} \sum_{j \in s} \Omega_{ij} y_i y_j \tag{1}$$

is a design-unbiased estimator of var$(\hat{\theta}_n)$, where Ω_{ij} depends on the joint inclusion probabilities $\pi_{ij} > 0$.

Under nonresponse, we define the response indicator variable for y_i as $\tilde{a}_i = 1$ if y_i is observed and $\tilde{a}_i = 0$ otherwise, $(i \in s)$. Conceptually, the definition of $\tilde{a}_i$ is extended to the entire population U. That is, we define a_i as $a_i = 1$ if y_i is observed when unit i is sampled and $a_i = 0$ otherwise, $i \in U$. Thus, $a_i = \tilde{a}_i$ for $i \in s$ and for every possible sample s.

Every method of estimation under missing data requires some assumptions about the population or response mechanism. There are two approaches to estimation of θ_N with missing data. Under the first, the population-model approach, a model for the distribution of the y_i is used without specifying the distribution of the a_i. Under the second, the quasi-randomization approach, the population y-values are treated as fixed and a model for the distribution of the a_i is assumed. The method of Särndal (1992) is based on the population-model approach that is used in this paper.

To impute for missing data under the population model approach, we assume a model for the population values y_i given x_i:

$$E_{\zeta}\left(y_i \mid x_i\right) = m\left(x_i; \beta_0\right) \tag{2}$$

for some p-dimensional vector β_0, where $m(x_i; \beta)$ is a known function of x_i for given β. The subscript ζ in (2) denotes that the reference distribution is the superpopulation model. We assume that the response mechanism is a missing at random response mechanism, i.e. the distribution of a_i depends only on x_i. Then, by (2),

$$E_{\zeta}\left[\sum_{i=1}^{N} a_i \left\{y_i - m\left(x_i, \beta_0\right)\right\} h\left(x_i\right) \mid X_N, A_N\right] = 0$$

holds for any $h(x)$, where $X_N = \{x_1, \ldots, x_N\}$ and $A_N = \{a_1, \ldots, a_N\}$.

Under deterministic imputation, suppose that we use

$$\hat{y}_i = m(x_i; \hat{\beta}) \tag{3}$$

as the imputed value for missing y_i based on the model (2), where $\hat{\beta}$ is the solution of estimating equations

$$\hat{U}(\beta) \equiv N^{-1} \sum_{i \in s} w_i a_i \{y_i - m(x_i; \beta)\} h(x_i; \beta) = 0 \tag{4}$$

for some p-dimensional vector $h(x_i, \beta)$, i.e. $\hat{U}(\hat{\beta}) = 0$. We assume that the solution $\hat{\beta}$ is unique. If the variance function $\mathrm{var}_\zeta(y_i \mid x_i)$ is specified as $\mathrm{var}_\zeta(y_i \mid x_i) = \sigma^2 q(x_i, \beta_0)$ for a known function $q(\cdot)$, then we choose $h(x_i, \beta) = \dot{m}(x_i, \beta)/q(x_i, \beta) \equiv h_i$, where $\dot{m}(x_i, \beta) = \partial m(x_i, \beta)/\partial \beta$. This choice is motivated by the quasilikelihood equations for generalized linear models (McCullagh & Nelder, 1989, Ch. 9). For commonly used population models, h_i^{T} is of the form $h_i^{\mathrm{T}} = (1, h_{1i}^{\mathrm{T}})$ in which case we have $\sum_{i \in s} w_i a_i \{y_i - m(x_i, \hat{\beta})\} = 0$ from (4). For example, the linear regression models (i) $m(x_i, \beta) = x_i^{\mathrm{T}} \beta$ with 1 as an element of x_i and $q(x_i, \beta) = 1$, (ii) $m(x_i, \beta) = x_i \beta$ and $q(x_i, \beta) = x_i$ and (iii) the logistic regression model with 1 as an element of x_i, all satisfy the above property of $h_i^{\mathrm{T}} = (1, h_{1i}^{\mathrm{T}})$.

Under deterministic imputation (3), the imputed estimator of the total θ_N can be written as

$$\hat{\theta}_{Id} = \sum_{i \in s} w_i \{a_i y_i + (1 - a_i) m(x_i; \hat{\beta})\}. \tag{5}$$

Theorem 1 below provides some asymptotic properties of the deterministically imputed estimator $\hat{\theta}_{Id}$ in (5).

THEOREM 1. *Assume that the finite population is a random sample from a superpopulation model* (2) *with finite* $(2+\delta)$*-th moments of* $(x_i, y_i, m_{i0}, \dot{m}_{i0}, h_{i0})$ *for some* $\delta > 0$*, where* $m_{i0} = m(x_i; \beta_0)$*,* $\dot{m}_{i0} = \dot{m}(x_i; \beta_0)$ *with* $\dot{m}(x_i; \beta) = \partial m(x_i; \beta)/\partial \beta$*,* $h_{i0} = h(x_i; \beta_0)$ *and* β_0 *satisfies* (2)*. Assume that the sampling mechanism and the response mechanism are ignorable under the model* (2)*. Assume that the solution to* (4) *is unique. Further assume that* (*i*) *the sampling design is such that, for any* z_i *with bounded* $2+\delta$ *moments,* $n\, var(N^{-1} \sum_{i \in s} w_i z_i \mid F_N) < K_{1z}$ *for some* $K_{1z} > 0$*, where* $F_N = \{z_1, \ldots, z_N\}$*,* (*ii*) *for each* i*,* $m(x_i; \beta)$ *and* $h(x_i; \beta)$ *are continuous functions of* β *in a compact set* B *containing* β_0 *as an interior point and* (*iii*) *for each* i*,* $m(x_i; \beta)$ *is differentiable with continuous partial derivative* $\dot{m}(x_i; \beta)$ *in a compact set containing* β_0*, and* $\sum_{i=1}^N a_i \dot{m}(x_i; \beta_0) h_{i0}^{\mathrm{T}}$ *is nonsingular. Then, the imputed estimator* (5) *satisfies*

$$n^{1/2} N^{-1} (\hat{\theta}_{Id} - \tilde{\theta}_{Id}) = o_p(1), \tag{6}$$

where

$$\tilde{\theta}_{Id} = \sum_{i \in s} w_i [m(x_i; \beta_0) + a_i \{1 + c^{\mathrm{T}} h(x_i; \beta_0)\} \{y_i - m(x_i; \beta_0)\}] \equiv \sum_{i \in s} w_i \eta_i \tag{7}$$

and

$$c = \left\{ \sum_{i=1}^N a_i \dot{m}\,(x_i; \beta_0)\, h_{i0}^{\mathrm{T}} \right\}^{-1} \sum_{i=1}^N (1 - a_i)\, \dot{m}\,(x_i; \beta_0). \tag{8}$$

Theorem 1 states that $\hat{\theta}_{Id}$ is asymptotically equivalent to $\tilde{\theta}_{Id}$ in (7). The reference distributions in (6) are the joint distribution of the superpopulation model (2) and the sampling mechanism, conditional on the realized values of (x_i, a_i) in the population. Proofs of Theorem 1 and of the asymptotic equivalence of $\mathrm{var}(\hat{\theta}_{Id})$ and $\mathrm{var}(\tilde{\theta}_{Id})$ may be found in the Appendix.

To study variance estimation, we use the reverse approach proposed by Fay (1992) and Shao & Steel (1999). In the reverse approach, the sample respondents s_R are regarded as a sample from the population of possible respondents U_R, and we have the representation: Population (U) $\rightarrow$ Responding Population (U_R) $\rightarrow$ Respondents (s_R). We apply the reverse approach to $\mathrm{var}(\tilde{\theta}_{Id})$ to get the following decomposition:

$$\mathrm{var}(\tilde{\theta}_{Id}) \equiv V_1 + V_2 + V_3,$$

where

$$V_1 \equiv E\left[E\left\{\mathrm{var}\left(\sum_{i\in s} w_i\eta_i \mid F_N, A_N\right) \mid A_N\right\}\right],$$

$$V_2 \equiv E\left[\mathrm{var}\left\{E\left(\sum_{i\in s} w_i\eta_i \mid F_N, A_N\right) \mid A_N\right\}\right],$$

$$V_3 \equiv \mathrm{var}\left[E\left\{E\left(\sum_{i\in s} w_i\eta_i \mid F_N, A_N\right) \mid A_N\right\}\right],$$

η_i is defined in (7), $F_N = \{y_1, \ldots, y_N\}$, and $A_N = \{a_1, \ldots, a_N\}$. The reference distribution in the conditional expectation given F_N and A_N is the sampling mechanism treating A_N as fixed. The reference distribution in the conditional expectation given A_N is over the superpopulation model (2) treating A_N as fixed. The marginal distribution with respect to A_N is over the unknown response mechanism. Now, by noting that $E(\sum_{i=1}^N \eta_i \mid A_N) = \sum_{i=1}^N m(x_i;\beta_0)$ does not depend on the response indicators a_i, it follows that $V_3 = 0$ and

$$\mathrm{var}(\tilde{\theta}_{Id}) \equiv V_1 + V_2. \tag{9}$$

To estimate $\mathrm{var}(\tilde{\theta}_{Id})$ in (9), we estimate the two terms, V_1 and V_2, separately. If β_0 were known, then η_i would be observed for all $i \in s$ and V_1 could then be estimated by applying the standard variance estimator formula (1) to the pseudo-values η_i. In practice, we replace β_0 by $\hat{\beta}$ to get a plug-in variance estimator,

$$\hat{V}_1 = \sum_{i\in s}\sum_{j\in s} \Omega_{ij}\hat{\eta}_i\hat{\eta}_j, \tag{10}$$

where Ω_{ij} is defined in (1) and $\hat{\eta}_i = \eta_i(\hat{\beta})$ are the pseudo-values for variance estimation. Using (7), the pseudo-value $\hat{\eta}_i$ is given by

$$\hat{\eta}_i = m(x_i;\hat{\beta}) + a_i(1 + \hat{c}^{\mathrm{T}}\hat{h}_i)\{y_i - m(x_i;\hat{\beta})\}, \tag{11}$$

where $\hat{h}_i = h(x_i, \hat{\beta})$ and

$$\hat{c} = \left\{\sum_{i\in s} w_i a_i \dot{m}(x_i;\hat{\beta})\hat{h}_i^{\mathrm{T}}\right\}^{-1} \sum_{i\in s} w_i\,(1 - a_i)\,\dot{m}(x_i;\hat{\beta}).$$

If $h(x_i, \beta) = h_i$ does not depend on β, then $\hat{h}_i = h_i$.

To estimate the second term V_2 in (9), we assume that $\mathrm{var}_\zeta(y_i \mid x_i) = \sigma^2 q(x_i;\beta_0)$. In this case, we have

$$V_2 = E\left\{\mathrm{var}\left(\sum_{i=1}^N \eta_i \mid \mathcal{A}_N\right)\right\} = \sigma^2 E\left\{\sum_{i=1}^N a_i\,(1 + c^{\mathrm{T}}h_i)^2\, q\,(x_i, \beta_0)\right\}.$$

Thus, a consistent estimator of V_2 is given by

$$\hat{V}_2 = \hat{\sigma}^2 \sum_{i \in s} w_i a_i (1 + \hat{c}^{\mathrm{T}} \hat{h}_i)^2 q(x_i, \hat{\beta}), \tag{12}$$

where $\hat{\sigma}^2$ is asymptotically a design-model unbiased estimator of σ^2. Alternatively, a more robust estimator of V_2 is given by

$$\tilde{V}_2 = \sum_{i \in s} w_i a_i (1 + \hat{c}^{\mathrm{T}} \hat{h}_i)^2 \{y_i - m(x_i, \hat{\beta})\}^2; \tag{13}$$

this is consistent for V_2 even if $\mathrm{var}_\zeta(y_i \mid x_i)$ is mis-specified, unlike $\hat{V}_2$.

Using the asymptotic equivalence of $\mathrm{var}(\hat{\theta}_{Id})$ and $\mathrm{var}(\tilde{\theta}_{Id})$, the variance estimator of $\hat{\theta}_{Id}$ is given by $\hat{V} = \hat{V}_1 + \hat{V}_2$ or $\tilde{V} = \hat{V}_1 + \tilde{V}_2$, where $\hat{V}_1$ is given by (10), $\hat{V}_2$ by (12) and $\tilde{V}_2$ by (13). The term V_2 is of smaller order than V_1 if the overall sampling rate is negligible (Shao & Steel, 1999) and in this case $\hat{V} \cong \hat{V}_1$. The actual implementation of the reverse approach under a negligible sampling rate can be easily carried out by inserting the pseudo-values $\hat{\eta}_i$ for y_i in the variance estimator (1) for the complete sample case. The pseudo-value (11) takes the form of $\hat{\eta}_i = \hat{y}_i + a_i g_i (y_i - \hat{y}_i)$, where $\hat{y}_i = m(x_i; \hat{\beta})$ and $g_i = 1 + \hat{c}^{\mathrm{T}} \hat{h}_i$ is a factor that is greater than one and accounts for the increase in the variance due to missingness. The choice of $g_i = 1$ in $\hat{\eta}_i$ leads to the naive variance estimator that treats imputed values as true values and leads to underestimation.

The g_i satisfy

$$\sum_{i \in s_R} w_i g_i \dot{m}(x_i; \hat{\beta}) = \sum_{i \in s} w_i \dot{m}(x_i; \hat{\beta}). \tag{14}$$

Note that

$$\frac{\partial}{\partial \beta} \tilde{\theta}_{Id}(\beta) = \sum_{i \in s} w_i \dot{m}(x_i; \beta) - \sum_{i \in s_R} w_i g_i \dot{m}(x_i; \beta).$$

Thus, condition (14) is essentially the condition that requires $\hat{\theta}_{Id}$ be independent of the estimated nuisance parameter β.

3. Illustrations

3·1. *Linear regression imputation*

We first consider the case of linear regression imputation, based on a model

$$E_\zeta(y_i \mid x_i) = x_i^{\mathrm{T}} \beta_0, \quad \mathrm{var}_\zeta(y_i \mid x_i) = \sigma^2, \tag{15}$$

where E_ζ and var_ζ in (15) respectively denote the expectation and variance with respect to the model and 1 is an element of x_i. The imputed estimator $\hat{\theta}_{Id}$ is given by (5) with $m(x_i, \hat{\beta}) = x_i^{\mathrm{T}} \hat{\beta}$, where $\hat{\beta} = (\sum_{i \in s} w_i a_i x_i x_i^{\mathrm{T}})^{-1} (\sum_{i \in s} w_i a_i x_i y_i)$. The pseudo-values for variance estimation can be written as $\hat{\eta}_i = x_i^{\mathrm{T}} \hat{\beta} + a_i (1 + \hat{c}^{\mathrm{T}} x_i)(y_i - x_i^{\mathrm{T}} \hat{\beta})$, where $\hat{c} = (\sum_{i \in s} w_i a_i x_i x_i^{\mathrm{T}})^{-1} \sum_{i \in s} (1 - a_i) w_i x_i$.

Using the fact that 1 is an element of x_i, we get

$$1 + \hat{c}^{\mathrm{T}} x_i = \left(\frac{\sum_{i \in s} w_i}{\sum_{i \in s} w_i a_i} \right) \{1 + (\bar{x}_n - \bar{x}_r)^{\mathrm{T}} S_{xxr}^{-} (x_i - \bar{x}_r)\}, \tag{16}$$

where

$$\bar{x}_n = \left(\sum_{i\in s} w_i\right)^{-1} \sum_{i\in s} w_i x_i,$$

$$\bar{x}_r = \left(\sum_{i\in s} w_i a_i\right)^{-1} \sum_{i\in s} w_i a_i x_i,$$

$$S_{xxr} = \left(\sum_{i\in s} w_i a_i\right)^{-1} \sum_{i\in s} w_i a_i (x_i - \bar{x}_r)(x_i - \bar{x}_r)^{\mathrm{T}},$$

and S_{xxr}^{-} is a generalized inverse of S_{xxr}. Note that $1 + \hat{c}^{\mathrm{T}} x_i$ is unique for any choice of generalized inverse. Here, $1 + \hat{c}^{\mathrm{T}} x_i$ is the inflation factor to account for the contribution of unit i in the deterministic imputation. The inflation factor $1 + \hat{c}^{\mathrm{T}} x_i$ in (16) satisfies $\sum_{i\in A_R} w_i(1 + \hat{c}^{\mathrm{T}} x_i)(1, x_i^{\mathrm{T}}) = \sum_{i\in s} w_i(1, x_i^{\mathrm{T}})$, which is a special case of (14).

3·2. *Ratio imputation*

Ratio imputation is based on a linear regression model given by

$$E_\zeta(y_i \mid x_i) = \beta_0 x_i, \quad \mathrm{var}_\zeta(y_i \mid x_i) = \sigma^2 x_i. \tag{17}$$

This model is often called the ratio model. The imputed estimator is given by (5) with $m(x_i, \hat{\beta}) = \hat{\beta} x_i$, where $\hat{\beta} = (\sum_{i\in s} w_i a_i x_i)^{-1} \sum_{i\in s} w_i a_i y_i$. Noting that $\hat{h}_i = h_i = 1$ under the ratio model (17), it is readily seen that

$$\hat{\eta}_i = \hat{\beta} x_i + a_i\,(1 + \hat{c})\,(y_i - \hat{\beta} x_i), \tag{18}$$

where $\hat{c} = (\sum_{i\in s} w_i a_i x_i)^{-1} \sum_{i\in s} w_i (1 - a_i) x_i$. Expression (18) agrees with the pseudo-value $\hat{\xi}_i$, equation (14) in Shao & Steel (1999), who obtained their (14) after several non-obvious steps, whereas our unified approach leads to (18) in a routine manner.

3·3. *Domain estimation*

The proposed linearization method can also be applied to the estimation of a subpopulation, also called domain, total $\theta_z = \sum_{i=1}^{N} z_i y_i$, where $z_i = 1$ if i belongs to the domain and $z_i = 0$ otherwise. The indicator variables z_i are observed for the sample units $i \in s$. The complete sample domain estimator can be written as $\hat{\theta}_z = \sum_{i\in s} w_i z_i y_i$. The imputed domain estimator under the linear regression model (15) is given by $\hat{\theta}_{I,z} = \sum_{i\in s} w_i z_i \{a_i y_i + (1 - a_i) x_i^{\mathrm{T}} \hat{\beta}\}$. The estimator $\hat{\theta}_{I,z}$ is design-model unbiased under (15). The pseudo-values $\hat{\eta}_{i,z}$ for variance estimation are given by $\hat{\eta}_{i,z} = z_i x_i^{\mathrm{T}} \hat{\beta} + a_i (z_i + \hat{c}_z^{\mathrm{T}} x_i)\,(y_i - x_i^{\mathrm{T}} \hat{\beta})$, where $\hat{c}_z = (\sum_{i\in s} w_i a_i x_i x_i^{\mathrm{T}})^{-1} \sum_{i\in s} w_i z_i (1 - a_i) x_i$. The pseudo-values $\hat{\eta}_{i,z}$ under the ratio model (17) are obtained similarly.

4. STOCHASTIC IMPUTATION

4·1. *Variance estimation*

The proposed method is now extended to stochastic imputation. Let y_i^* be the imputed value of a missing y_i under a stochastic imputation method such that $E_I(y_i^*) = m(x_i; \hat{\beta}) \equiv \hat{m}_i$, the imputed value from a deterministic imputation, and E_I denotes the conditional expectation over the imputation mechanism. We define the stochastically imputed estimator as

$$\hat{\theta}_I = \sum_{i \in s} w_i \{a_i y_i + (1 - a_i) y_i^*\}, \tag{19}$$

where the imputed values, y_i^*, are independently generated. Note that $E_I(\hat{\theta}_I) = \hat{\theta}_{Id}$, where $\hat{\theta}_{Id}$ is given by (5). The total variance of $\hat{\theta}_I$ in (19) is decomposed as

$$\mathrm{var}(\hat{\theta}_I) = \mathrm{var}(\hat{\theta}_{Id}) + \mathrm{var}_I(\hat{\theta}_I - \hat{\theta}_{Id}), \tag{20}$$

where var_I denotes conditional variance over the imputation mechanism. The first term, $V(\hat{\theta}_{Id})$, in (20) can be estimated from the linearization method by $\hat{V}$ or $\tilde{V}$ given in § 3. To estimate the second term, $\mathrm{var}(\hat{\theta}_I - \hat{\theta}_{Id})$, in (20), note that $E_I(\hat{\theta}_I - \hat{\theta}_{Id}) = 0$ and $\mathrm{var}(\hat{\theta}_I - \hat{\theta}_{Id}) = E\{\mathrm{var}_I(\hat{\theta}_I - \hat{\theta}_{Id})\}$. Since $\hat{\theta}_I - \hat{\theta}_{Id} = \sum_{i \in s} w_i (1 - a_i)(y_i^* - \hat{m}_i)$, we have

$$\mathrm{var}_I(\hat{\theta}_I - \hat{\theta}_{Id}) = \sum_{i \in s} w_i^2 (1 - a_i)(y_i^* - \hat{m}_i)^2 \equiv V^*, \tag{21}$$

noting that the y_i^* are independently generated. The total variance of $\hat{\theta}_I$ is now estimated by $\hat{V} + V^*$ or $\tilde{V} + V^*$, where V^* is given by (21).

4·2. *Hot deck imputation*

The proposed approach to variance estimation can be applied to estimate the variance of the imputed estimator under hot deck imputation within imputation cells, where the imputed values are randomly selected with replacement from the respondents in the same cell. The underlying population model for hot deck imputation within cells is the cell mean model; see Kim & Fuller (2004). Under hot deck imputation, $E_I(y_i^*) = (\sum_{i \in s_g} w_i h_i a_i)^{-1} (\sum_{i \in s_g} w_i h_i a_i y_i) \equiv \bar{y}_{rg}$ for the imputed value y_i^* in cell g, where s_g is the set of sample indices in cell g and the donors are selected with probability proportional to $w_i h_i$. The choice of $h_i = 1$ leads to the weighted hot deck imputation considered in Rao & Shao (1992) and $h_i = w_i^{-1}$ leads to unweighted hot deck imputation. The pseudo-values in cell g for variance estimation can be written as $\hat{\eta}_i = \bar{y}_{rg} + a_i(1 + \hat{c}_g)(y_i - \bar{y}_{rg})$, where $\hat{c}_g = (\sum_{i \in s_g} w_i h_i a_i)^{-1} \sum_{i \in s_g} w_i (1 - a_i)$. The pseudo-values are applied to (10) to get $\hat{V}_1$ and V_2 is estimated by $\tilde{V}_2$ in (13) with $m(x_i; \hat{\beta})$ replaced by $\bar{y}_{rg}$. For the hot deck imputation, the total variance of $\hat{\theta}_I$ is now estimated by $\hat{V}_1 + \tilde{V}_2 + V^*$, where V^* is given by (21) with $\hat{m}_i = m(x_i; \hat{\beta})$.

4·3. *Multiple imputation*

The proposed approach to variance estimation can also be used to estimate the variance of the imputed estimator under the multiple imputation approach of Rubin (1987). In multiple imputation, M imputed values, $y_i^{*(1)}, \ldots, y_i^{*(M)}$, are generated independently for each missing item y_i, and the imputed estimator of θ_N is obtained as $\hat{\theta}_{MI} = M^{-1} \sum_{k=1}^{M} \hat{\theta}_I^{(k)}$, where $\hat{\theta}_I^{(k)} = \sum_{i \in s} w_i \{a_i y_i + (1 - a_i) y_i^{*(k)}\}$. Again, $E_I(\hat{\theta}_{MI}) = \hat{\theta}_{Id}$ provided $E_I\{y_i^{*(k)}\} = \hat{m}_i$. Hence, the variance decomposition (20) still holds with $\hat{\theta}_I$ changed to $\hat{\theta}_{MI}$. In Rubin (1987), the first term, $\mathrm{var}(\hat{\theta}_{Id})$, in (20) is estimated by $W_M + B_M$, where W_M is the average of the M naive variance estimators of the $\hat{\theta}_I^{(k)}$ and $B_M = (M - 1)^{-1} \sum_{k=1}^{M} (\hat{\theta}_I^{(k)} - \hat{\theta}_{MI})^2$. The second term $\mathrm{var}_I(\hat{\theta}_I - \hat{\theta}_{Id})$ in (20) for $\hat{\theta}_I = \hat{\theta}_{MI}$ is unbiasedly estimated by $M^{-1} B_M$. Rubin's variance estimator is theoretically justified when

$$\mathrm{var}(\hat{\theta}_{Id}) = \mathrm{var}(\hat{\theta}_n) + \mathrm{var}(\hat{\theta}_{Id} - \hat{\theta}_n), \tag{22}$$

where $\hat{\theta}_n$ is the complete sample estimator of θ_N. Meng (1994) called assumption (22) the 'congeniality' assumption. Fay (1992) and Kim et al. (2006) discussed situations where the congeniality assumption does not hold and thus $W_M + B_M$ can be biased for var$(\hat{\theta}_{Id})$. For example, the congeniality condition (22) is not satisfied for domain estimation when the domains are not specified at the imputation stage. On the other hand, the proposed linearization approach in § 3 can be used to estimate var$(\hat{\theta}_{Id})$ or var$(\hat{\theta}_{I,z})$ without the congeniality condition (22). Thus, a valid estimator of var$(\hat{\theta}_{MI})$ is given by $\hat{V}_{MI} = \hat{V} + M^{-1}B_M$ or $\tilde{V}_{MI} = \tilde{V} + M^{-1}B_M$, where $\hat{V}$ and $\tilde{V}$ are consistent estimators of var$(\hat{\theta}_{Id})$.

4·4. *Binary response*

We now consider the case of a binary response y_i taking the values 1 or 0 and obeying the logistic linear regression model $y_i \mid x_i \sim \text{Ber}\{m_i = m(x_i; \beta_0)\}$, where $\log\{m_i/(1-m_i)\} = x_i^{\mathrm{T}}\beta_0$ and 1 is an element of x_i. In this case, $E_\zeta(y_i \mid x_i) = m(x_i, \beta_0)$ and $\text{var}_\zeta(y_i \mid x_i) = q(x_i, \beta_0) = m_i(1-m_i)$. The estimator $\hat{\beta}$ is obtained iteratively from the estimating equations

$$\hat{U}(\beta) = \sum_{i \in s} w_i a_i \{y_i - m(x_i, \beta)\} x_i = 0, \tag{23}$$

noting that $\dot{m}(x_i, \beta_0) = m_i(1-m_i)x_i$, $V_\zeta(y_i \mid x_i) = m_i(1-m_i)$ and $h_i = x_i$.

Stochastic hot deck imputation for the binary response case is implemented by imputing $y_i^* = 1$ for missing y_i with probability $\hat{m}_i = m(x_i, \hat{\beta})$ and $y_i^* = 0$ with probability $1 - \hat{m}_i$. This method satisfies the condition $E_I(y_i^*) = \hat{m}_i$. The imputed estimator $\hat{\theta}_I$ is given by (19). For variance estimation, it follows from (11) and the above expressions for $\dot{m}(x_i, \beta)$ and h_i that the pseudo-value $\hat{\eta}_i$ is given by

$$\hat{\eta}_i = \hat{m}_i + a_i(1 + \hat{c}^{\mathrm{T}}x_i)(y_i - \hat{m}_i), \tag{24}$$

where

$$\hat{c} = \left\{\sum_{i \in s} w_i a_i \hat{m}_i(1-\hat{m}_i)x_i x_i^{\mathrm{T}}\right\}^{-1} \sum_{i \in s} w_i(1-a_i)\hat{m}_i(1-\hat{m}_i)x_i.$$

Using the $\hat{\eta}_i$ given by (24), we obtain $\hat{V}$ or $\tilde{V}$, and V^* is given by (21). Thus, the total variance of $\hat{\theta}_I$ is given by $\hat{V} + V^*$ or $\tilde{V} + V^*$.

The component V^* due to stochastic imputation can be eliminated by using fractional imputation (Kim & Fuller, 2004) with $M = 2$ fractions as follows: impute $y_i^* = 1$ for missing y_i with fractional weight $\hat{m}_i = m(x_i, \hat{\beta})$ and $y_i^* = 0$ with fractional weight $1 - \hat{m}_i$. The data file under fractional imputation will report real values 1 and 0 with associated fractions $\hat{m}_i$ and $1 - \hat{m}_i$ for a missing y_i, unlike the data file under deterministic imputation reporting $y_i^* = \hat{m}_i$. The imputed estimator $\hat{\theta}_{FI}$ of the total θ_N under fractional imputation reduces to $\hat{\theta}_{Id}$ that uses deterministic imputation $y_i^* = \hat{m}_i$ because $1(\hat{m}_i) + 0(1-\hat{m}_i) = \hat{m}_i$. Hence, the stochastic component V^* is eliminated for $\hat{\theta}_{Id}$ and the total variance of $\hat{\theta}_{Id}$ is estimated by $\hat{V}$ or $\tilde{V}$ using the pseudo-values $\hat{\eta}_i$ given by (24).

For the estimation of a domain total, $\theta_{N,z}$, the imputed estimator is given by $\hat{\theta}_{FI,z} = \sum_{i \in s} w_i z_i \{a_i y_i + (1-a_i)\hat{m}_i\}$. The linearization variance estimator under fractional imputation is obtained from the pseudo-values $\hat{\eta}_{i,z} = z_i\hat{m}_i + a_i(z_i + \hat{c}_z^{\mathrm{T}}x_i)(y_i - \hat{m}_i)$, where

$$\hat{c}_z = \left\{\sum_{i \in s} a_i w_i \hat{m}_i(1-\hat{m}_i)x_i x_i^{\mathrm{T}}\right\}^{-1} \sum_{i \in s}(1-a_i) w_i \hat{m}_i(1-\hat{m}_i) z_i x_i.$$

Note that $E_I(y_i^*) = \hat{m}_i$, as stipulated in § 4. The imputed estimator of a domain mean, $\bar{\theta}_{N,z} = (\sum_{i=1}^N z_i)^{-1}\theta_{N,z}$, is given by $(\sum_{i\in s} w_i z_i)^{-1}\hat{\theta}_{FI,z}$. The associated linearization variance estimator is obtained from the pseudo-values

$$\tilde{\eta}_{i,z} = \left(\sum_{i\in s} w_i z_i\right)^{-1} \left\{\hat{\eta}_{i,z} - \left(\sum_{i\in s} w_i z_i\right)^{-1} \hat{\theta}_{FI,z}\right\}. \tag{25}$$

5. Composite imputation

Consider three items, x_i, y_i and z_i, for each unit i. Assume that z_i is always observed while x_i and y_i are subject to missingness. The parameter of interest is the population total, θ_N, of y and it is estimated by $\hat{\theta} = \sum_{i\in s} w_i y_i$ under complete response, as before. Under the presence of missing data, the sample s can be written as $s = s_{\mathrm{RR}} \cup s_{\mathrm{RM}} \cup s_{\mathrm{MR}} \cup s_{\mathrm{MM}}$, where s_{RR}, s_{RM}, s_{MR} and s_{MM} respectively denote the sample elements for which both x_i and y_i are observed, only y_i is missing, only x_i is missing, and both x_i and y_i are missing. Also, define $s_{\mathrm{R+}} = s_{\mathrm{RR}} \cup s_{\mathrm{RM}}$ and $s_{\mathrm{M+}} = s_{\mathrm{MR}} \cup s_{\mathrm{MM}}$.

Under the above set-up, we consider composite imputation based on the linear regression model given by

$$E_\zeta(y_i \mid x_i, z_i) = \beta_{y|x} x_i \qquad E_\zeta(x_i \mid z_i) = \beta_{x|z} z_i. \tag{26}$$

The nuisance parameters, $\beta_{y|x}$ and $\beta_{x|z}$, in (26) are estimated from the estimating equations

$$\hat{U}_1(\beta_{y|x}) \equiv \sum_{i\in s_{\mathrm{RR}}} w_i(y_i - \beta_{y|x} x_i) = 0$$

and

$$\hat{U}_2(\beta_{x|z}) \equiv \sum_{i\in s_{\mathrm{R+}}} w_i(x_i - \beta_{x|z} z_i) = 0,$$

leading to

$$\hat{\beta}_{y|x} = \left(\sum_{i\in s_{\mathrm{RR}}} w_i x_i\right)^{-1} \sum_{i\in s_{\mathrm{RR}}} w_i y_i, \quad \hat{\beta}_{x|z} = \left(\sum_{i\in s_{\mathrm{R+}}} w_i z_i\right)^{-1} \sum_{i\in s_{\mathrm{R+}}} w_i x_i.$$

The estimators $\hat{\beta}_{y|x}$ and $\hat{\beta}_{x|z}$ are model unbiased for $\beta_{y|x}$ and $\beta_{x|z}$, respectively. Based on the estimators $\hat{\beta}_{y|x}$ and $\hat{\beta}_{x|z}$, the proposed composite imputation involves the following two steps. (i) If $i \in s_{\mathrm{RM}}$, impute y_i by $\hat{y}_i = \hat{\beta}_{y|x} x_i$ where $\hat{\beta}_{y|x} = (\sum_{i\in s_{\mathrm{RR}}} w_i x_i)^{-1} \sum_{i\in s_{\mathrm{RR}}} w_i y_i$. (ii) If $i \in s_{\mathrm{MM}}$, impute y_i by $\tilde{y}_i = \hat{\beta}_{y|x}\hat{x}_i$ where $\hat{x}_i = \hat{\beta}_{x|z} z_i$ and $\hat{\beta}_{x|z} = (\sum_{i\in s_{\mathrm{R+}}} w_i z_i)^{-1} \sum_{i\in s_{\mathrm{R+}}} w_i x_i$, where $s_{\mathrm{R+}} = s_{\mathrm{RR}} \cup s_{\mathrm{RM}}$. Thus, $\tilde{y}_i = \hat{\beta}_{y|x}\hat{\beta}_{x|z} z_i$.

The imputed estimator under the above composite imputation can be written as

$$\hat{\theta}_I = \hat{\theta}_I(\hat{\beta}_{y|x}, \hat{\beta}_{x|z}) = \sum_{i\in s_{+\mathrm{R}}} w_i y_i + \sum_{i\in s_{\mathrm{RM}}} w_i x_i \hat{\beta}_{y|x} + \sum_{i\in s_{\mathrm{MM}}} w_i z_i \hat{\beta}_{y|x}\hat{\beta}_{x|z}, \tag{27}$$

where $s_{+\mathrm{R}} = s_{\mathrm{RR}} \cup s_{\mathrm{MR}}$. The imputed estimator (27) is design-model unbiased under the model (26).

To apply the linearization approach under the proposed composite imputation, using the linearization formula

$$\hat{\theta}_I(\hat{\beta}) \cong \hat{\theta}_I(\beta_0) - E\left\{\frac{\partial \hat{\theta}_I(\beta_0)}{\partial \beta^{\mathrm{T}}}\right\}\left[E\left\{\frac{\partial \hat{U}(\beta_0)}{\partial \beta^{\mathrm{T}}}\right\}\right]^{-1} \hat{U}(\beta_0), \tag{28}$$

where $\beta = (\beta_{y|x}, \beta_{x|z})^{\mathrm{T}}$ and $\hat{U} = (\hat{U}_1, \hat{U}_2)^{\mathrm{T}}$, we have

$$\begin{aligned}\hat{\theta}_I \cong & \sum_{i \in s_{\mathrm{RR}}} w_i\{x_i\beta_{y|x} + c_{1x}(1 + c_{2z})(y_i - x_i\beta_{y|x}) + \beta_{y|x}c_{2z}(x_i - \beta_{x|z}z_i)\} \\ & + \sum_{i \in s_{\mathrm{RM}}} w_i\beta_{y|x}\{x_i + c_{2z}(x_i - \beta_{x|z}z_i)\} + \sum_{i \in s_{\mathrm{MM}}} w_i z_i \beta_{y|x}\beta_{x|z},\end{aligned}$$

where

$$c_{1x} = \frac{E\left(\sum_{i \in s_{\mathrm{R}+}} w_i x_i\right)}{E\left(\sum_{i \in s_{\mathrm{RR}}} w_i x_i\right)}, \quad c_{2z} = \frac{E\left(\sum_{i \in s_{\mathrm{MM}}} w_i z_i\right)}{E\left(\sum_{i \in s_{\mathrm{R}+}} w_i z_i\right)}.$$

Therefore, the pseudo-values for variance estimation, based on the linearization approach, are given by

$$\hat{\eta}_i = \begin{cases} x_i\hat{\beta}_{y|x} + \hat{c}_{1x}(1 + \hat{c}_{2z})(y_i - x_i\hat{\beta}_{y|x}) + \hat{\beta}_{y|x}\hat{c}_{2z}(x_i - \hat{\beta}_{x|z}z_i), & i \in s_{\mathrm{RR}}, \\ \beta_{y|x}\{x_i + \hat{c}_{2z}(x_i - \hat{\beta}_{x|z}z_i)\}, & i \in s_{\mathrm{RM}}, \\ z_i\hat{\beta}_{y|x}\hat{\beta}_{x|x}, & i \in s_{\mathrm{MM}}, \end{cases}$$

where $\hat{c}_{1x} = \sum_{i \in s_{\mathrm{R}+}} w_i x_i / \sum_{i \in s_{\mathrm{RR}}} w_i x_i$, $\hat{c}_{2z} = \sum_{i \in s_{\mathrm{MM}}} w_i z_i / \sum_{i \in s_{\mathrm{R}+}} w_i z_i$.

The resulting linearization variance estimator is approximately design-model unbiased because both the design-model expectation of $\hat{U}_1(\beta_{y|x})$ and that of $\hat{U}_2(\beta_{x|z})$ are zero under the model (26).

6. Simulation study

We performed a simulation study to validate the proposed linearization method. In the simulation, $B = 5000$ finite populations, of size $N = 10\,000$, were first generated independently from an infinite population specified by $x_i \sim N(3, 1)$, $y_i \mid x_i \sim \mathrm{Ber}\{m_i = m(x_i, \beta_0)\}$, $u_i \mid (x_i, y_i) \sim N(2 + 0{\cdot}5x_i, 0{\cdot}75)$ and $z_i \mid (x_i, y_i, u_i) \sim \mathrm{Ber}\,(0{\cdot}4)$, where $\log\{m_i/(1 - m_i)\} = 0{\cdot}5x_i - 2$ with $\beta_0 = (-2, 0{\cdot}5)^{\mathrm{T}}$. Among the four variables (x_i, y_i, z_i, u_i), only y_i is subject to missingness. A sample of size $n = 100$ was then selected from each of the 5000 simulated finite populations.

The simulation set-up employed a $2 \times 3 \times 3$ factorial structure with three factors. Factor 1 refers to the sampling mechanism with two levels, where level 1 denotes simple random sampling and level 2 refers to sampling with probability proportional to size u_i. Factor 2 corresponds to the response mechanism given by $a_i \sim \mathrm{Ber}(\pi_i)$, where $\mathrm{logit}(\pi_i) = \phi_0 + \phi_1(x_i - 3) + \phi_2|x_i - 3|$. The three levels of factor 2 are specified by (a) $\phi_1 = 0, \phi_2 = 0$, (b) $\phi_1 = 1, \phi_2 = 0$ and (c) $\phi_1 = 0, \phi_2 = 1$. In each case, ϕ_0 is determined to achieve 70% overall response rate. Factor 2 refers to the imputation mechanism with three levels specified by (a) multiple imputation of Rubin (1987) with $M = 5$, (b) fractional imputation using $y_i^* = 1$ with fractional weight $\hat{m}_i$ and $y_i^* = 0$ with fractional weight $1 - \hat{m}_i$, where $\hat{m}_i = m(x_i, \hat{\beta})$ and $\hat{\beta}$ is obtained iteratively from

Table 1. *Relative bias of linearized variance estimators*

Parameter	Sampling design	Response mechanism	% Relative bias FI	% Relative bias HD
Population mean	SRS	a	2·74	1·36
		b	0·09	−0·85
		c	−2·68	−0·61
	PPS	a	3·96	2·11
		b	5·58	5·36
		c	3·97	5·08
Domain mean	SRS	a	2·11	0·45
		b	1·59	−0·02
		c	−3·49	−0·63
	PPS	a	2·11	1·83
		b	3·52	2·96
		c	0·56	1·00

FI, fractional imputation; HD, hot deck; SRS, simple random sampling; PPS, probability proportional to size sampling; a, b and c correspond to the cases ($\phi_1 = 0, \phi_2 = 0$), ($\phi_1 = 1, \phi_2 = 0$) and ($\phi_1 = 1, \phi_2 = 1$), respectively.

the estimating equations (23), and (c) hot deck imputation based on y_i^* generated from Ber$(\hat{m}_i)$. In the case of multiple imputation, a vector β^* was first generated from

$$\beta^* \sim N\left[\hat{\beta},\ \left\{\sum_{i\in s} a_i \hat{m}_i\,(1-\hat{m}_i)\,(1,x_i)\,(1,x_i)^{\mathrm{T}}\right\}^{-1}\right].$$

From the generated β^*, an imputed value y_i^* for missing y_i was then generated as $y_i^* \sim \mathrm{Ber}(m_i^*)$, where $m_i^* = m(x_i, \beta^*)$. This process was repeated $M = 5$ times independently to generate the imputed values $y_i^{*(1)}, \ldots, y_i^{*(5)}$.

The linearization variance estimator is given by $\hat{V}$ for fractional imputation and $\hat{V} + V^*$ for hot deck imputation, where the pseudo-value $\hat{\eta}_i$ is given by (24) in the case of the population mean and by (25) for the domain mean. Table 1 presents the simulated values of the relative bias of the variance estimator under the specified factorial structure. The size of simulation error for the values reported in Table 1 is about 0·2%. Table 1 shows that the absolute values of the relative bias are all small, less than 4%. Table 2 compares the relative bias of Rubin's variance estimator $W_M + (1 + M^{-1})B_M$, and the linearization variance estimator $\hat{V} + M^{-1}B_M$, in the case of multiple imputation. The size of simulation error for the relative bias of linearization variance estimator reported in Table 2 is about 0·2% but it is larger for the relative bias of Rubin's variance estimator because it has fewer degrees of freedom. For example, assuming 16 degrees of freedom, the simulation error is about 0·5%. Table 2 shows that the values of absolute relative biases of the linearization variance estimators are small in all the cases, less than 7%. On the other hand, Rubin's variance estimator leads to large values of absolute relative bias in the case of domain mean, ranging from 28% to 34%. The congeniality condition (22) is not satisfied here because the domains are not specified at the design stage, thus leading to large relative bias.

7. Concluding remarks

For simplicity, we have focused on imputed estimators based only on the design weights and derived a linearization variance estimator based on the pseudo-values $\hat{\eta}_i$ given by (11). However, our results can readily be extended to imputed estimators based on calibration weights, $\tilde{w}_i$, satisfying the calibration constraints $\sum_{i\in s} \tilde{w}_i x_{Ci} = \sum_{i\in U} x_{Ci}$, where the x_{Ci} are some calibration

Table 2. *Relative bias of Rubin's variance estimator and the linearization variance estimators for multiple imputation*

Parameter	Sampling design	Response mechanism	% Relative bias	
			R	L
Population mean	SRS	a	1·07	2·90
		b	−0·29	1·42
		c	−3·96	−2·09
	PPS	a	−2·52	4·95
		b	0·50	7·00
		c	−2·48	5·45
Domain mean	SRS	a	34·25	2·37
		b	31·08	2·28
		c	27·55	−3·41
	PPS	a	27·97	2·63
		b	32·41	4·61
		c	27·93	1·75

R, Rubin's variance estimator; L, linearization variance estimator; SRS, simple random sampling; PPS, probability proportional to size sampling; a, b and c correspond to the cases $(\phi_1 = 0, \phi_2 = 0)$, $(\phi_1 = 1, \phi_2 = 0)$ and $(\phi_1 = 1, \phi_2 = 1)$, respectively.

variables with known total $\sum_{i \in U} x_{Ci}$ obtained from external sources. The linearization variance estimator corresponding to $\hat{V}_1$, given by (10), can be obtained from standard formulae for calibration estimators in the complete-data case by substituting $\hat{\eta}_i$ for y_i. The component corresponding to $\hat{V}_2$ is obtained by changing the design weight w_i in (13) to the corresponding calibration weight $\tilde{w}_i$. In the case of stochastic imputation, the additional component is given by (21) with the design weight w_i changed to the calibration weight $\tilde{w}_i$. Davison & Sardy (2007) heuristically obtained linearization variance estimators in the case of calibration estimators under deterministic linear regression imputation for stratified random sampling.

We have derived the linearization (28) under the population model approach, but it can also be justified under the quasi-randomization approach as long as $E\{\hat{U}(\beta_0)\} = 0$ holds under the quasi-randomization approach. Variance estimation under the quasi-randomization approach is under investigation.

The proposed linearization method of variance estimation may need modification when $\hat{\theta}_n$ is not linear in the y_i, such as the quantiles. Moreover, the proposed method, as it stands, does not cover nearest neighbour imputation (Chen & Shao, 2001).

ACKNOWLEDGEMENT

The work of the first author was supported by a Korea Science and Engineering Foundation grant. The work of the second author was supported by a research grant from the Natural Sciences and Engineering Research Council of Canada. The authors wish to thank the referees and the associate editor for their constructive comments.

APPENDIX

Proofs

Proof of Theorem 1. First, define $U(\beta) = E\{\hat{U}(\beta) \mid X_N, A_N\}$ where $X_N = \{x_1, \ldots, x_N\}$ and $A_N = \{a_1, \ldots, a_N\}$. Here, the expectation is taken with respect to the joint distribution of the superpopulation model (2) and the sampling mechanism. Note that β_0 satisfies $U(\beta) = 0$. To prove Theorem 1, we need the following two lemmas.

LEMMA 1. *Under conditions* 1 *and* 2 *of Theorem* 1, *we have*

$$\hat{\beta} - \beta_0 = o_p(1). \tag{A1}$$

Proof of Lemma 1. The pointwise convergence of $\hat{U}(\beta)$ to $U(\beta)$ in probability follows from condition 1. By condition 2 and the compactness of B, the convergence is uniform, i.e.

$$\sup_{\beta \in \mathcal{B}} |\hat{U}(\beta) - U(\beta)| = o_p(1). \tag{A2}$$

By the continuity of $U(\beta)$ in B, given any $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\text{pr}(|\hat{\beta} - \beta_0| > \epsilon) \leqslant \text{pr}\{|U(\hat{\beta}) - U(\beta_0)| > \delta\}$$

holds. Further,

$$\begin{aligned} \text{pr}\{|U(\hat{\beta}) - U(\beta_0)| > \delta\} &\leqslant \text{pr}\{|\hat{U}(\hat{\beta}) - U(\hat{\beta})| > \delta/2\} + \text{pr}\{|\hat{U}(\hat{\beta}) - U(\beta_0)| > \delta/2\} \\ &\leqslant \text{pr}\left\{\sup_{\beta \in \mathcal{B}} |\hat{U}(\beta) - U(\beta)| > \delta/2\right\} + 0 \end{aligned}$$

because $\hat{U}(\hat{\beta}) = U(\beta_0) = 0$. Thus, using (A2), the consistency property (A1) follows.

LEMMA 2. *Under conditions* 1–3 *of Theorem* 1,

$$\hat{\beta} - \beta_0 = -\{H(\beta_0)\}^{-1} \hat{U}(\beta_0) + o_p(n^{-1/2}), \tag{A3}$$

where $H(\beta) = \partial U(\beta)/\partial\beta$. □

Proof of Lemma 2. Since (A1) holds, we can apply the mean value theorem

$$\hat{U}(\hat{\beta}) - \hat{U}(\beta_0) = \hat{H}(\beta^*)(\hat{\beta} - \beta_0), \tag{A4}$$

where $\hat{H}(\beta) = \partial\hat{U}(\beta)/\partial\beta$ and β^* is a point between $\hat{\beta}$ and β_0. By conditions 1 and 2, we can establish the pointwise convergence of $\hat{H}(\beta)$ to $H(\beta)$ in probability.

Using the fact that $\mathcal{B}$ is a compact set and $\hat{H}(\beta)$ and $H(\beta)$ are uniformly continuous in B, we can apply the argument used for (A2) to get the uniform convergence of $\hat{H}(\beta)$ to $H(\beta)$ in probability. Thus, by the continuity of $H(\beta)$, we have

$$\hat{H}(\beta^*) = H(\beta_0) + o_p(1). \tag{A5}$$

Thus, by (A5) and $\hat{U}(\hat{\beta}) = 0$, (A4) reduces to

$$-n^{1/2}\hat{U}(\beta_0) = n^{1/2}H(\beta_0)(\hat{\beta} - \beta_0) + o_p\left(n^{1/2}\|\hat{\beta} - \beta_0\|\right). \tag{A6}$$

By the Cauchy–Schwarz inequality and (A6),

$$\begin{aligned} n^{1/2}\|\hat{\beta} - \beta_0\| &\leqslant \|H^{-1}(\beta_0)\|\|n^{1/2}H(\beta_0)(\hat{\beta} - \beta_0)\| \\ &= O_p(1) + o_p\left(n^{1/2}\|\hat{\beta} - \beta_0\|\right), \end{aligned}$$

which implies $n^{1/2}$-consistency of $\hat{\beta}$. Thus, (A6) reduces to

$$-n^{1/2}\hat{U}(\beta_0) = n^{1/2}H(\beta_0)(\hat{\beta} - \beta_0) + o_p(1). \tag{A7}$$

Hence, (A3) follows from (A7). □

Proof of (6). To prove (6), first define $\hat{\theta}_I(\beta) = \sum_{i\in s} w_i\{a_i y_i + (1-a_i)m(x_i;\beta)\}$ and consider the class of estimators $\hat{\theta}(\beta,k) = \hat{\theta}_I(\beta) + Nk^{\mathrm{T}}\hat{U}(\beta)$, where $\hat{U}(\beta)$ is given by (4). Note that

$$\hat{\theta}\,(\beta,k) = \sum_{i\in s} w_i[m(x_i;\beta) + a_i\{1 + k^{\mathrm{T}}h(x_i;\beta)\}\{y_i - m(x_i;\beta)\}],$$

which can be expressed as $\hat{\theta}\,(\beta,k) = \sum_{i\in s} w_i\eta_i(\beta,k)$ for some $\eta_i(\beta,k)$ that is a known function of (x_i, y_i, a_i) up to β and k.

We now find a particular choice of k, say k^*, such that

$$\hat{\theta}\,(\hat{\beta},k^*) = \hat{\theta}\,(\beta_0,k^*) + o_p(n^{-1/2}N). \tag{A8}$$

Thus, if condition (A8) holds, then the effect of estimating β can be safely ignored by choosing $k = k^*$. Since $\hat{\theta}\,(\hat{\beta},k) = \hat{\theta}_{Id}$ for all p-dimensional vectors k, (A8) implies that

$$\hat{\theta}_{Id} = \hat{\theta}\,(\beta_0,k^*) + o_p(n^{-1/2}N). \tag{A9}$$

To find k^* that satisfies (A8), using the theory of Randles (1982), noting that $\surd n(\hat{\beta}-\beta_0) = O_p(1)$ by (A3), the asymptotic equivalence (A8) holds if $E\{\partial\hat{\theta}(\beta,k^*)/\partial\beta \mid \beta = \beta_0\} = 0$, which in turn holds if

$$k^* = -N^{-1}\left[E\left\{\left.\frac{\partial\hat{U}\,(\beta)^{\mathrm{T}}}{\partial\beta}\right|_{\beta=\beta_0}\right\}\right]^{-1} E\left\{\left.\frac{\partial\hat{\theta}_I(\beta)}{\partial\beta}\right|_{\beta=\beta_0}\right\} = c, \tag{A10}$$

where $E\,(\cdot)$ denotes the design-model expectation and c is given by (8). The result (6) now follows from (A9) and (A10), noting that $\hat{\theta}(\beta_0,k^*) = \tilde{\theta}_{Id}$. □

Asymptotic equivalence of var$(\hat{\theta}_{Id})$ *and var*$(\tilde{\theta}_{Id})$. The variance of $\hat{\theta}_{Id}$ is asymptotically equivalent to the variance of $\tilde{\theta}_{Id}$ if

$$\mathrm{var}(\hat{\theta}_{Id}) = \mathrm{var}(\tilde{\theta}_{Id}) + o\left(\frac{N^2}{n}\right). \tag{A11}$$

To prove (A11), we need to establish the following steps.

Step 1.

$$E\left\{n(\hat{\beta}-\beta_0)^2\right\} = O(1). \tag{A12}$$

Step 2.

$$\hat{\theta}_{Id} = \hat{\theta}_{Id}\,(\beta_0) + \hat{Q}\,(\beta_0)\,(\hat{\beta}-\beta_0) + Z_n, \tag{A13}$$

where $\hat{Q}(\beta) = \partial\hat{\theta}_{Id}(\beta)/\partial\beta$ and $E(Z_n^2) = O(n^{-2}N^2)$.

Step 3.

$$\hat{\theta}_{Id} = \tilde{\theta}_{Id} + W_n + Z_n, \tag{A14}$$

where $E(W_n^2) = O(n^{-2}N^2)$ and Z_n is defined in (A13).

Step 4. $E\{(\hat{\theta}_{Id} - \tilde{\theta}_{Id})\} = O(n^{-1}N)$ and $E\{(\hat{\theta}_{Id} - \tilde{\theta}_{Id})^2\} = o(n^{-1}N^2)$.

Proof of Step 1. By the mean value theorem,

$$\begin{aligned} 0 = \hat{U}(\hat{\beta}) &= \hat{U}\,(\beta_0) + \hat{H}\,(\beta^*)\,(\hat{\beta}-\beta_0) \\ &= \hat{U}\,(\beta_0) + H(\beta_0)\,(\hat{\beta}-\beta_0) + \{\hat{C}(\beta_0) + \hat{D}(\beta^*)\}(\hat{\beta}-\beta_0), \end{aligned} \tag{A15}$$

where $\hat{H}(\beta) = \partial \hat{U}(\beta)/\partial\beta$, $H(\beta) = \partial U(\beta)/\partial\beta$, $\hat{C}(\beta) = \hat{H}(\beta) - H(\beta)$ and $\hat{D}(\beta^*) = \hat{H}(\beta^*) - \hat{H}(\beta_0)$. By the Cauchy–Schwarz inequality,

$$E[\{\hat{C}(\beta_0) + \hat{D}(\beta^*)\}(\hat{\beta} - \beta_0)] \leqslant [E\{\hat{C}(\beta_0) + \hat{D}(\beta^*)\}^2 E(\hat{\beta} - \beta_0)^2]^{1/2}.$$

Note that $E\{\hat{C}^2(\beta_0)\} = o(1)$. Write

$$\hat{D}(\beta) = N^{-1} \sum_{i \in S} w_i a_i \{\dot{u}(x_i, y_i; \beta) - \dot{u}(x_i, y_i; \beta_0)\},$$

where $\dot{u}(x_i, y_i; \beta) = \partial u(x_i, y_i; \beta)/\partial\beta$ and $u(x_i, y_i; \beta) = \{y_i - m(x_i, \beta)\}h(x_i; \beta)$. By the continuity of $\dot{u}(x_i, y_i; \beta)$, given any $\epsilon > 0$, there exists a $\delta > 0$ such that

$$E\{\hat{D}^2(\beta)\} \leqslant N^{-1} \sum_{i=1}^{N} a_i \{\dot{u}(x_i, y_i; \beta) - \dot{u}(x_i, y_i; \beta_0)\}^2 \leqslant \epsilon/2 \tag{A16}$$

holds for all β such that $|\beta - \beta_0| \leqslant \delta$. Also, by the uniform continuity of $\dot{u}(x_i, y_i; \beta)$ in B,

$$E\left\{\sup_{\beta \in \mathcal{B}} \hat{D}^2(\beta)\right\} \leqslant K_D \tag{A17}$$

for some K_D. Then, by $\beta^* - \beta_0 = o_p(1)$ for the $\delta > 0$ satisfying (A16), we can find $n_0 = n_0(\delta)$ such that

$$\text{pr}(|\beta^* - \beta_0| > \delta) \leqslant \epsilon/2K_D \tag{A18}$$

for all $n \geqslant n_0$. Thus, for all $n \geqslant n_0$,

$$E\{\hat{D}^2(\beta^*)\} \leqslant |E\{\hat{D}^2(\beta^*) \mid \beta^* \in B(\beta_0, \delta)\}| + K_D \text{pr}\{\beta^* \notin B(\beta_0, \delta)\} \leqslant \epsilon, \tag{A19}$$

by (A16), (A17) and (A18), where $B(\beta_0, \delta) = \{\beta; |\beta - \beta_0| < \delta\}$. Thus, we have

$$E[\{\hat{C}(\beta_0) + \hat{D}(\beta^*)\}^2] = o(1)$$

and, by (A15),

$$\hat{\beta} - \beta_0 = -[H(\beta_0) + o(1)]^{-1}\hat{U}(\beta_0).$$

Now, noting that $H(\beta_0) = O(1)$ and $E\{\hat{U}(\beta_0)^2\} = O(n^{-1})$, we get (A12).

Proof of Step 2. By the second-order Taylor expansion,

$$\hat{\theta}_{Id} = \hat{\theta}_{Id}(\beta_0) + \hat{Q}(\beta_0)(\hat{\beta} - \beta_0) + 0{\cdot}5(\hat{\beta} - \beta_0)^{\mathrm{T}}\hat{A}(\beta^*)(\hat{\beta} - \beta_0),$$

where $\hat{Q}(\beta) = \partial\hat{\theta}_{Id}(\beta)/\partial\beta$ and $\hat{A}(\beta) = \partial^2\hat{\theta}_I(\beta)/(\partial\beta\partial\beta^{\mathrm{T}})$. Now, letting

$$Z_n = 0{\cdot}5(\hat{\beta} - \beta_0)^{\mathrm{T}}\hat{A}(\beta^*)(\hat{\beta} - \beta_0),$$

we have

$$\begin{aligned} Z_n &= 0{\cdot}5(\hat{\beta} - \beta_0)^{\mathrm{T}}\hat{A}(\beta_0)(\hat{\beta} - \beta_0) + 0{\cdot}5(\hat{\beta} - \beta_0)^{\mathrm{T}}\{\hat{A}(\beta^*) - \hat{A}(\beta_0)\}(\hat{\beta} - \beta_0) \\ &\equiv (\hat{\beta} - \beta_0)^{\mathrm{T}}(Z_{n1} + Z_{n2})(\hat{\beta} - \beta_0). \end{aligned}$$

Using the same argument for establishing (A19), it can be shown that $E(Z_{n1}^2) = O(N^2)$ and $E(Z_{n2}^2) = o(N^2)$. Now, using (A12), we have $E(Z_n) = O(n^{-1}N)$ and (A13). $\square$

Proof of Step 3. Using $\hat{\beta} - \beta_0 = -\hat{H}^{-1}(\beta^*)\hat{U}(\beta_0)$, we have

$$\begin{aligned} W_n &= \hat{\theta}_{Id}(\beta_0) + \hat{Q}(\beta_0)(\hat{\beta} - \beta_0) - \tilde{\theta}_{Id} \\ &= -\{\hat{Q}(\beta_0)\hat{H}^{-1}(\beta_0) - Q(\beta_0)H^{-1}(\beta_0)\}\hat{U}(\beta_0) - \hat{Q}(\beta_0)\{\hat{H}^{-1}(\beta^*) - \hat{H}^{-1}(\beta_0)\}\hat{U}(\beta_0) \\ &\equiv W_{n1} + W_{n2}. \end{aligned}$$

Thus, using standard arguments, $E(W_{n1}^2) = O(n^{-1}N^2) \times O(n^{-1}) = O(n^{-2}N^2)$. Also, using the same argument for (A19), $E(W_{n2}^2) = o(n^{-2}N^2)$. Thus, (A14) follows. □

Proof of Step 4. Directly follows from Steps 2 and 3. □

REFERENCES

CHEN, J. & SHAO, J. (2001). Jackknife variance estimation for nearest-neighbor imputation. *J. Am. Statist. Assoc.* **96**, 260–69.

DAVISON, A. C. & SARDY, S. (2007). Resampling variance estimation in surveys with missing data. *J. Offic. Statist.* **23**, 371–86.

DEVILLE, J.C. & SÄRNDAL, C. E. (1994). Variance estimation for the regression imputed Horvitz-Thompson estimator. *J. Offic. Statist.* **10**, 381–94.

FAY, R. E. (1992). When are inferences from multiple imputation valid? In *Proc. Survey Res. Meth. Sect., Am. Statist. Assoc.,* pp. 227–32. Washington, DC: American Statistical Association.

FULLER, W. A. & KIM, J. K. (2005). Hot deck imputation under response model. *Survey Methodol.* **31**, 139–49.

HAZIZA, D. (2009). Imputation and inference in the presence of missing data. In *Handbook of Statistics*. Vol. 29, *Sample Surveys: Theory, Methods and Inference,* Ed. D. Pfeffermann & C.R. Rao, pp. 215–46. Amsterdam: Elsevier BV.

KIM, J. K. & FULLER, W. A. (2004). Inference procedures for hot deck imputation. *Biometrika* **91**, 559–78.

KIM, J. K., BRICK, M. J., FULLER, W. A. & KALTON, G. (2006). On the bias of the multiple imputation variance estimator in survey sampling. *J. R. Statist. Soc.* B **68**, 509–21.

MCCULLAGH, P. & NELDER, J. (1989). *Generalized Linear Models*. London: Chapman & Hall.

MENG, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statist. Sci.* **9**, 538–73.

RANDLES, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *Ann. Statist.* **10**, 462–74.

RAO, J. N. K. & SHAO, J. (1992). Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* **79**, 811–22.

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

SÄRNDAL, C. E. (1992). Methods for estimating the precision of survey estimates when imputation is used. *Survey Methodol.* **18**, 241–52.

SHAO, J. & STEEL, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *J. Am. Statist. Assoc.* **94**, 254–65.

[*Received March* 2008. *Revised February* 2009]