Linearization Variance Estimators for Survey Data: Some Recent Work

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

A. Demnati, Social Survey Methods Division, Statistics Canada, Ottawa, Canada[1]
J. N. K. Rao, School of Mathematics and Statistics, Carleton University, Ottawa, Canada[2]

Abstract

In survey sampling, Taylor linearization is often used to obtain variance estimators of calibration estimators of totals and nonlinear finite population parameters. It is generally applicable to any sampling design, but it can lead to multiple variance estimators that are asymptotically design unbiased under repeated sampling. The choice among the variance estimators requires other considerations such as (i) approximate unbiasedness for the model variance of the estimator under an assumed model, and (ii) validity under a conditional repeated sampling framework. Demnati and Rao (2004) proposed a new approach to deriving Taylor linearization variance estimators that leads directly to a unique variance estimator that satisfies the above considerations for general designs. Demnati and Rao (2002) considered the case of missing responses when adjustment for complete nonresponse and imputation for item nonresponse are used. Demnati and Rao (2003) extended the work to deal with longitudinal surveys which lead to dependent observations and to multiple weights on the same element. They considered a variety of longitudinal sampling designs, covering panel surveys, household panel surveys as well as rotating surveys. Demnati and Rao (2005) 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 fraction is small, 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. In this paper, we give a brief account of the Demnati-Rao method for variance estimation. We also present simulation results on total variance estimation and some extensions.

Keywords: Calibration, model parameters, total variance.

1. Introduction

Taylor linearization is a popular method of variance estimation for complex statistics such as ratio and regression estimators and logistic regression coefficient estimators. It is generally applicable to any sampling design that permits unbiased variance estimation for linear estimators, unlike a resampling method such as the jackknife, and it is computationally simpler than the latter method. 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, and (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 $\vartheta_L = N^2(n^{-1} - N^{-1})s_e^2$ does not track the conditional MSE 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 an auxiliary variable x, s_e^2 is the sample variance of the residuals $e_k = y_k - (\bar{y}/\bar{x})x_k$ and (n, N) denote the sample and population sizes. By linearizing the jackknife variance estimator, ϑ_J, we obtain a different linearization variance estimator, $\vartheta_{JL} = (\bar{X}/\bar{x})^2 \vartheta_L$, which 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_L. Valliant (1993) obtained ϑ_{JL} for the post-stratified estimator and conducted a simulation study to demonstrate that both ϑ_J and ϑ_{JL} possess good conditional properties given the estimated post-strata counts. Särndal, Swensson and Wretman (1989) showed that ϑ_{JL} is both asymptotically design unbiased and asymptotically model unbiased in the sense of $E_m(\vartheta_{JL}) = V_m(\hat{Y}_R)$, where E_m denotes model expectation and $V_m(\hat{Y}_R)$ 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 $V_m(y_k) = \sigma^2 x_k$, $\sigma^2 > 0$. Thus, ϑ_{JL} is a good choice from either the design-based or the model-based perspectives.

Demnati and Rao (2004) proposed a new 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 (2002) extended their method to the case of missing responses when adjustments for complete nonresponse and imputation for item nonresponse based on smooth functions of observed values, in particular ratio imputation, are used. Demnati and Rao (2003) extended the work to deal with longitudinal surveys which lead to dependent observations and to multiple weights on the same element. They considered a variety of longitudinal sampling designs, covering panel surveys, household panel surveys as well as rotating surveys. Demnati and Rao (2005) 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, 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.

In this paper, we give a brief account of the Demnati-Rao (DR) method for variance estimation. In section 2, we review the DR method for total variance estimation. We apply the method to the ratio estimator and provide simulation results on the performance of DR variance estimator. In section 3, we extend the results to estimators of model parameters defined as solutions to weighted estimating equations. Results in section 3 are extended to the case of multiple weight adjustments in section 4.

2. Demnati-Rao Linearization Method

We start with a general formulation of the Demnati and Rao (2004) approach to deriving Taylor linearization variance estimators. This formulation will cover both finite population (or census) parameters, θ_N, and model parameters, θ, under an assumed super-population model. An estimator, $\hat{\theta}$, based on a probability sample, s, drawn from a finite population P of size N is used to estimate both θ_N and θ. However, variance estimators associated with θ_N and θ are different. In the latter case, we estimate the total variance $V(\hat{\theta}) = E_m V_p(\hat{\theta}) + V_m E_p(\hat{\theta})$, while the design variance $V_p(\hat{\theta})$ is estimated in the former case, where E_m and V_m denote model expectation and model variance and E_p and V_p denote design expectation and design variance, respectively.

Let $\boldsymbol{d}_k = (d_{1k},...,d_{gk})^T$ be a $g \times 1$ vector of random weights and $\boldsymbol{u}_k = (u_{1k},...,u_{gk})^T$ be a $g \times 1$ vector of constants for $k = 1,...,N$. Let $\hat{U} = \sum \boldsymbol{u}_k^T \boldsymbol{d}_k$ be a linear estimator and using an operator notation let $\vartheta(\boldsymbol{u})$ denote the estimator of variance of $\hat{U}$, where $\sum$ denotes summation over all elements k in P. We write $\hat{\theta}$ as $f(\boldsymbol{A}_d)$, where $\boldsymbol{A}_d$ is $g \times N$ matrix with k^{th} column $\boldsymbol{d}_k$. The choice of $\boldsymbol{A}_d$ depends on the random processes involved. For example, suppose $\theta_N = \sum y_k$ is the population total and we use a ratio estimator $\hat{\theta} = X(\sum y_k d_k)/(\sum x_k d_k) = X\hat{R}$, where $d_k = a_k / \pi_k$ are the Horvitz-Thompson weights with $a_k = 1$ if element k is in the sample s, $a_k = 0$ otherwise and π_k are the inclusion probabilities. In this case, $g = 1$, $\boldsymbol{d}_k = d_k$, $\boldsymbol{A}_d = (d_1,...,d_N)^T$, $f(\boldsymbol{A}_d) = X(\sum y_k d_k)/(\sum x_k d_k) = X\hat{Y}/\hat{X}$, and $\vartheta(u) = \vartheta_p(u)$ is the design-based variance estimator of the total $\hat{U} = \sum u_k d_k$:

$$\vartheta_p(u) = \sum\sum u_k u_t d_{kt}(1-\omega_{kt})/\omega_{kt}, \qquad (2.1)$$

where $d_{kt} = a_k a_t / \pi_{kt}$, $k \neq t$, $d_{kk} = d_k$, $\omega_{kt} = \pi_k \pi_t / \pi_{kt}$, and π_{kt} are the joint inclusion probabilities.

Suppose, on the other hand, we are interested in the model parameter $\theta = \sum E_m(y_k)$ under a superpopulation model on y_k. In this case, $g = 2$, $d_{1k} = d_k$, $d_{2k} = d_k y_k$, and $f(\boldsymbol{A}_d) = X(\sum d_{2k})/(\sum x_k d_{1k})$. Further, the estimator of total variance of $\hat{U}$ is

$$\vartheta(\boldsymbol{u}) = \sum\sum \boldsymbol{u}_k^T \operatorname{cov}(\boldsymbol{d}_k, \boldsymbol{d}_t)\boldsymbol{u}_t, \qquad (2.2)$$

where $\operatorname{cov}(\boldsymbol{d}_k, \boldsymbol{d}_t)$ is an estimator of total variance of $\boldsymbol{d}_k$ and $\boldsymbol{d}_t$. In the above case of $\boldsymbol{d}_k = d_k \boldsymbol{v}_k$, where $\boldsymbol{v}_k = (1, y_k)^T$, we have

$$\begin{aligned}\operatorname{cov}(\boldsymbol{d}_k, \boldsymbol{d}_t) &= d_{kt}\begin{bmatrix} 0 & 0 \\ 0 & \operatorname{cov}_m(y_k, y_t)\end{bmatrix} \\ &+ d_{kt}\frac{(1-\omega_{kt})}{\omega_{kt}}\boldsymbol{v}_k \boldsymbol{v}_t^T.\end{aligned} \qquad (2.3)$$

In (2.3), $\operatorname{cov}_m(y_k, y_t)$ is an estimator of the covariance of y_k and y_t under the assumed model. When the model covariance of y_k and y_t is zero, $\operatorname{cov}_m(y_k, y_t)$ is taken as zero.

The DR linearization variance estimator of $\hat{\boldsymbol{\theta}} = f(\boldsymbol{A}_b)$ is simply given by

$$\vartheta_{DR}(\hat{\theta}) = \vartheta(\boldsymbol{z}), \qquad (2.4)$$

where $\vartheta(\boldsymbol{z})$ is obtained from $\vartheta(\boldsymbol{u})$ by replacing $\boldsymbol{u}_k$ by $\boldsymbol{z}_k = \partial f(\boldsymbol{A}_b)/\partial \boldsymbol{b}_k |_{A_b = A_d}$, where $\boldsymbol{A}_b$ is a $g \times N$ matrix of arbitrary real numbers with k^{th} column $\boldsymbol{b}_k = (b_{1k},\dots,b_{gk})^T$. In the case of θ_N, we use $\vartheta_p(\boldsymbol{z})$ in (2.4).

Application to Ratio Estimator

For the ratio estimator $\hat{\theta}$ and the finite population total θ_N, we have $f(\boldsymbol{A}_b) = X(\sum y_k b_k)/(\sum x_k b_k)$ and hence

$$\boldsymbol{z}_k = z_k = (X/\hat{X})(y_k - \hat{R}\,x_k) = (X/\hat{X})e_k. \quad (2.5)$$

The DR variance estimation is then given by $\vartheta_p(z)$. Similarly, for the model parameter $\theta = \sum E_m(y_k)$, we have $f(\boldsymbol{A}_b) = X(\sum b_{2k})/(\sum x_k b_{1k})$ and

$$\boldsymbol{z}_k = \begin{pmatrix} z_{1k} \\ z_{2k}\end{pmatrix} = \frac{X}{\hat{X}}\begin{pmatrix} -\hat{R}\,x_k \\ 1\end{pmatrix}. \qquad (2.6)$$

Substituting $\boldsymbol{z}_k$ in (2.6) for $\boldsymbol{u}_k$ in (2.2), we get

$$\begin{aligned}\vartheta_{DR}(\hat{\theta}) &= \sum\sum d_{kt} z_{k;m} z_{t;m} \operatorname{cov}_m(y_k, y_t) \\ &+ \sum\sum d_{kt} z_{k;s} z_{t;s}(1-\omega_{kt})/\omega_{kt} \\ &\equiv \vartheta_m + \vartheta_s\end{aligned} \qquad (2.7)$$

where $z_{k;m} = z_{2k} = X/\hat{X}$ and $z_{k;s} = \boldsymbol{z}_k^T \boldsymbol{v}_k = z_{1k} + z_{2k} y_k = (X/\hat{X})(y_k - \hat{R}x_k)$.

Note that the first component, ϑ_m, corresponds to the model while the second component, ϑ_s, corresponds to the sampling design.

Now consider the special case of simple random sampling without replacement (SRS). For this special case,

$$\vartheta_s = \frac{N^2}{n}\left(1 - \frac{n}{N}\right)\left(\frac{X}{\hat{X}}\right)^2 s_e^2 = \vartheta_p(z), \qquad (2.8)$$

where $s_e^2 = \sum a_k e_k^2/(n-1)$. Further, under the ratio model

$$E_m(y_k) = \beta x_k,\ Cov_m(y_k, y_t) = 0,\ k \neq t, \quad (2.9)$$

$\theta = \beta X$, and the model variance of y_k, $V_m(y_k) = E_m(y_k - \beta x_k)^2$ is estimated robustly by $\vartheta_m(y_k) = e_k^2$. The model component ϑ_m of (2.7) reduces to

$$\vartheta_m = \frac{N}{n}\left(\frac{X}{\hat{X}}\right)^2 (n-1)s_e^2. \qquad (2.10)$$

Note that ϑ_m remains valid under misspecification of $V_m(y_k)$. Now combining (2.8) and (2.10), we get

$$\vartheta_{DR}(\hat{\theta}) = \frac{N^2}{n}\left(\frac{X}{\hat{X}}\right)^2 \frac{N-1}{N} s_e^2. \qquad (2.11)$$

It is interesting to note that the "g-weight" appears automatically in $\vartheta_{DR}(\hat{\theta})$, and that the finite population correction $1 - n/N$ is absent in $\vartheta_{DR}(\hat{\theta})$. The DR approach leads to a unique choice of variance estimator that preserves the g-factors automatically.

It is customary to ignore the factor $(X/\hat{X})^2$ and use

$$\vartheta_{cus}(\hat{\theta}) = \frac{N(N-1)}{n} s_e^2 . \quad (2.12)$$

A mixed version which includes the g-factor in the sampling component only, i.e., uses ϑ_s given by (2.8), is given by

$$\vartheta_{mix}(\hat{\theta}) = \frac{N}{n} s_e^2 \left[(N-n)\left(\frac{X}{\hat{X}}\right)^2 + (n-1) \right]. \quad (2.13)$$

Here, $V_m E_p(\hat{\theta}) \approx V_m(Y) = \sum E_m(y_k - \beta x_k)^2$ is estimated by $\vartheta_m^* = \sum d_k e_k^2 = (N/n)(n-1)s_e^2$ which is used in place of ϑ_m in (2.7).

We conducted a small simulation study to examine the performances of different variance estimators, both unconditionally and conditionally on $\hat{X}$. We first generated $R = 2{,}000$ finite populations $\{y_k\}$ each of size $N = 393$, from the ratio model

$$y_k = 2x_k + x_k^{1/2}\varepsilon_k , \quad (2.14)$$

with ε_k are independent observations generated from a $N(0,1)$, where the fixed x_k are the "number of beds" for the Hospitals population studied in Valliant *et al*. (2000, p.424-427). One simple random sample of specified size n is drawn from each generated population. Our parameter of interest is $\theta = \beta X = 2X$, and the simulated total MSE of the ratio estimator $\hat{\theta} = X(\bar{y}/\bar{x})$ is calculated as $M(\hat{\theta}) = R^{-1}\sum_{r=1}^{2000}(\hat{\theta}_r - \theta)^2$, where $\hat{\theta}_r$ is the value of $\hat{\theta}$ for the r^{th} simulated sample and $(\bar{y}, \bar{x})$ are the sample means. We calculated the total variance estimate $\vartheta_{DR}(\hat{\theta})$, and its components ϑ_s and ϑ_m from each simulated sample and their averages $\bar{\vartheta}_{DR}$, $\bar{\vartheta}_s$, and $\bar{\vartheta}_m$. Figure 1 gives a plot of the average of variance estimates $\bar{\vartheta}_{DR}$ and $\bar{\vartheta}_s$, and the simulated MSE for $n = 20, 40, \cdots, 380, 393$. In the case of $n = N$, $\bar{\vartheta}_s = 0$. It is seen from Figure 1, that $\bar{\vartheta}_{DR}$ tracks $M(\hat{\theta})$ very well, whereas the use of $\bar{\vartheta}_s$ leads to severe underestimation as the sample size, n, increases.

To examine the conditional performance of the variance estimators under simple random sampling given $\hat{X} = N\bar{x}$, we conducted another simulation study, similar to Royall and Cumberland (1981) for inference on $\theta = \sum E_m(y_k)$ using model (2.14). We generated $R = 20{,}000$ finite populations $\{y_k\}$ each of size $N = 393$ from (2.14) using the number of beds as x_k and from each population we then selected one simple random sample of size $n = 100$. We arranged the 20,000 samples in ascending order of $\bar{x}$-values and then grouped them into 20 groups each of size 1,000 such that the first group, G_1, contained 1,000 samples with the smallest $\bar{x}$-values, the next group, G_2, contained the next 1,000 smallest $\bar{x}$-values, and so on to get $G_1,...,G_{20}$. For each of the 20 groups so formed, we calculated the average values of the ratio estimates $\hat{\theta} = X(\bar{y}/\bar{x})$ and the expansion estimates $N\bar{y}$, and the resulting conditional relative bias (CRB) in estimating $\theta = 2X$; see Figure 2. It is clear from Figure 2 that $N\bar{y}$ is conditionally biased unlike $\hat{\theta}$: Negative CRB (-14%) for G_1 increasing to positive CRB (+14%) for G_{20}. Note that both $N\bar{y}$ and $\hat{\theta}$ are unconditionally unbiased for θ. The conditional bias of $\hat{\theta}$ and $N\bar{y}$ in estimating θ is similar to the conditional bias in estimating $\theta_N = Y$, observed by Royall and Cumberland (1981).

We also calculated the conditional MSE of $\hat{\theta}$ and the associated CRB of the variance estimates ϑ_{DR}, ϑ_{cus} and ϑ_{mix} based on the average values of ϑ_{DR}, ϑ_{cus} and ϑ_{mix} in each group; see Figure 3. It is evident from Figure 3 that CRB of ϑ_{cus} ranges from -28% to 20% across the groups whereas ϑ_{DR} exhibits no such trend and its CRB is less than 5% in absolute value except for G_6 and G_{20}. Also, the CRB of ϑ_{mix} is largely negative and below that of ϑ_{DR} for the first half of the groups and above for the second half, but ϑ_{mix} exhibits no visible trends unlike ϑ_{cus}. Figure 4 reports the conditional coverage rates (CCR) of normal theory confidence intervals based on ϑ_{DR}, ϑ_{cus}, ϑ_{mix} and ϑ_s (ignoring the component ϑ_m) for nominal level of 95%. As expected, the use of ϑ_s leads to severe undercoverage because the sampling fraction, 100/393, is significant. On the other hand, CCR associated with ϑ_{DR} is closer to nominal level across groups, while ϑ_{cus} exhibits a trend across groups with CCR ranging from 91% to 97%. Further, CCR associated with ϑ_{mix} is below that of ϑ_{DR} for the first half of the groups and above for the second half.

3. Estimating Equations

Suppose the supperpopulation model on the responses y_k is specified by a generalized linear model with mean $E_m(y_k) = \mu_k(\boldsymbol{\theta}) = h(\boldsymbol{z}_k^T \boldsymbol{\theta})$, where $\boldsymbol{z}_k$ is a $p \times 1$ vector of explanatory variables and $h(.)$ is a "link" function. The model parameter of interest is $\boldsymbol{\theta}$. For example, the choice $h(a) = a$ gives linear regression model and $h(a) = e^a/(1+e^a)$ leads to the logistic regression model for binary responses y_k. We define census estimating equations (CEE) as $\boldsymbol{l}_N(\boldsymbol{\theta}) = \sum \boldsymbol{l}_k(\boldsymbol{\theta}) = \boldsymbol{0}$ with $E_m \boldsymbol{l}_k(\boldsymbol{\theta}) = \boldsymbol{0}$, and the solution to CEE gives the census parameter $\boldsymbol{\theta}_N$. We have $\boldsymbol{l}_k(\boldsymbol{\theta}) = \boldsymbol{z}_k(y_k - \mu_k(\boldsymbol{\theta}))$ for linear and logistic regression models.

We use a general class of calibration estimators with weights $w_k = d_k F(\boldsymbol{x}_k^T \hat{\boldsymbol{\alpha}})$, where the vector parameter $\boldsymbol{\alpha}$ is determined by solving a set of calibration constraints

$$\sum d_k F(\boldsymbol{x}_k^T \boldsymbol{\alpha}) \boldsymbol{x}_k = \boldsymbol{X}, \tag{3.1}$$

where $\boldsymbol{X}$ is the known total of a $q \times 1$ vector of calibration variables $\boldsymbol{x}_k$. For example, the choice $F(a) = 1 + a$ gives GREG weights and $F(a) = e^a$ leads to raking ratio weights. We use the calibration weights to estimate the CEE. The calibration weighted estimating equations are given by

$$\hat{\boldsymbol{l}}(\boldsymbol{\theta}) = \sum d_k F(\boldsymbol{x}_k^T \hat{\boldsymbol{\alpha}}) \boldsymbol{l}_k(\boldsymbol{\theta}) = \boldsymbol{0}. \tag{3.2}$$

The solution to (3.2) gives the calibration weighted estimator $\hat{\boldsymbol{\theta}}$ of both $\boldsymbol{\theta}_N$ and $\boldsymbol{\theta}$. It is approximately design unbiased for $\boldsymbol{\theta}_N$ and design-model unbiased for $\boldsymbol{\theta}$, i.e., $E_p(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}_N$ and $E_m E_p(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$. We focus here on total variance estimation associated with $\boldsymbol{\theta}$. Demnati and Rao (2004) studied the case of $\boldsymbol{\theta}_N$ under the general class of calibration weights, while Demnati and Rao (2005) studied the case of model parameter $\boldsymbol{\theta}$ under GREG weights. It follows from (3.2) that $\hat{\boldsymbol{\theta}}$ is of the form $\boldsymbol{f}(\boldsymbol{A}_d)$ with $\boldsymbol{d}_k = (d_k, d_k \boldsymbol{l}_k^T(\boldsymbol{\theta}))^T$, where $\boldsymbol{f}(\boldsymbol{A}_d)$ is a $p \times 1$ vector and $\boldsymbol{A}_d$ is a $(p+1) \times N$ matrix with k^{th} column $\boldsymbol{d}_k$. Following the implicit differentiation method of Demnati and Rao (2004), $\boldsymbol{Z}_k = \partial \boldsymbol{f}(\boldsymbol{A}_b)/\partial \boldsymbol{b}_k |_{\boldsymbol{A}_b = \boldsymbol{A}_d}$ is evaluated as

$$\boldsymbol{Z}_k^T = [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})]^{-1} F(\boldsymbol{x}_k^T \hat{\boldsymbol{\alpha}})(-\hat{\boldsymbol{B}}^T(\boldsymbol{l}, \hat{\boldsymbol{\alpha}})\boldsymbol{x}_k, \boldsymbol{I}_p), \tag{3.3}$$

with

$$\hat{\boldsymbol{B}}(\boldsymbol{l}, \boldsymbol{\alpha}) = [\sum d_k f(\boldsymbol{x}_k^T \boldsymbol{\alpha}) \boldsymbol{x}_k \boldsymbol{x}_k^T]^{-1} \sum d_k f(\boldsymbol{x}_k^T \boldsymbol{\alpha}) \boldsymbol{x}_k \boldsymbol{l}_k^T(\hat{\boldsymbol{\theta}}), \tag{3.4}$$

$$\hat{\boldsymbol{J}}(\boldsymbol{\theta}) = -\sum d_k F(\boldsymbol{x}_k^T \hat{\boldsymbol{\alpha}})(\partial \boldsymbol{l}_k(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^T), \tag{3.5}$$

$\boldsymbol{I}_p$ is the identity matrix and $f(a) = \partial F(a)/\partial a$. The DR linearization variance estimator of $\hat{\boldsymbol{\theta}}$ is obtained from (2.2) and (2.3) by replacing $\boldsymbol{u}_k^T$ by the $p \times (p+1)$ matrix $\boldsymbol{Z}_k^T$, $\boldsymbol{v}_k$ by $(1, \boldsymbol{l}_k^T(\boldsymbol{\theta}))^T$ and $\text{cov}_m(y_k, y_t)$ by an estimator of the $p \times p$ covariance matrix of $\boldsymbol{l}_k(\boldsymbol{\theta})$ under the assumed model. After simplification, we get

$$\vartheta_{DR}(\hat{\boldsymbol{\theta}}) = \vartheta_m + \vartheta_s, \tag{3.6}$$

where ϑ_s is the sampling estimated covariance matrix given by

$$\vartheta_s = [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})]^{-1} \sum\sum d_{kt} F(\boldsymbol{x}_k^T \hat{\boldsymbol{\alpha}}) F(\boldsymbol{x}_t^T \hat{\boldsymbol{\alpha}}) [(1-\omega_{kt})/\omega_{kt}] \boldsymbol{e}_k^*(\hat{\boldsymbol{\theta}}) \boldsymbol{e}_t^*(\hat{\boldsymbol{\theta}}) [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})]^{-1\mathrm{T}} \tag{3.7}$$

with

$$\boldsymbol{e}_k^*(\hat{\boldsymbol{\theta}}) = \boldsymbol{l}_k(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{B}}^T(\boldsymbol{l}, \hat{\boldsymbol{\alpha}}) \boldsymbol{x}_k. \tag{3.8}$$

The model estimated covariance matrix, ϑ_m, depends on the assumed model covariance structure. If $Cov_m(\boldsymbol{l}_k(\boldsymbol{\theta}), \boldsymbol{l}_t^T(\boldsymbol{\theta})) = \boldsymbol{0}$ for $k \neq t$, and $\boldsymbol{V}_m(\boldsymbol{l}_k(\boldsymbol{\theta})) = E_m(\boldsymbol{l}_k(\boldsymbol{\theta}) \boldsymbol{l}_k^T(\boldsymbol{\theta}))$ is estimated robustly by $\boldsymbol{l}_k(\hat{\boldsymbol{\theta}}) \boldsymbol{l}_k^T(\hat{\boldsymbol{\theta}})$, then the model estimated covariance matrix, ϑ_m, reduces to

$$\vartheta_m = [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})]^{-1} \sum d_k F^2(\boldsymbol{x}_k^T \hat{\boldsymbol{\alpha}}) \boldsymbol{l}_k(\hat{\boldsymbol{\theta}}) \boldsymbol{l}_k^T(\hat{\boldsymbol{\theta}}) [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})]^{-1\mathrm{T}}. \tag{3.9}$$

Note that for the linear regression and logistic regression models, $Cov_m(\boldsymbol{l}_k(\boldsymbol{\theta}), \boldsymbol{l}_t^T(\boldsymbol{\theta})) = \boldsymbol{0}$ for $k \neq t$ if the y_k's are uncorrelated under the model, noting that $\boldsymbol{l}_k(\boldsymbol{\theta}) = \boldsymbol{z}_k(y_k - \mu_k(\boldsymbol{\theta}))$.

We conducted a small simulation study to examine the unconditional (design-model) performance of the calibration weighted estimator $\hat{\boldsymbol{\theta}}$ of the model parameter $\boldsymbol{\theta}$ under a Poisson regression model and simple post-stratification adjustment. In particular, we compared the efficiency of $\hat{\boldsymbol{\theta}}$ relative to $\tilde{\boldsymbol{\theta}}$ using only design weights in the weighted estimating equations (3.2): $\tilde{\boldsymbol{l}}(\boldsymbol{\theta}) = \sum d_k \boldsymbol{l}_k(\boldsymbol{\theta}) = \boldsymbol{0}$. We also examined the unconditional performance of the variance estimators $\vartheta_{DR}(\hat{\boldsymbol{\theta}})$ and $\vartheta_{DR}(\tilde{\boldsymbol{\theta}})$ in tracking the total variances of $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$, respectively. Note that $\vartheta_{DR}(\hat{\boldsymbol{\theta}})$ is given by (3.6) and $\vartheta_{DR}(\tilde{\boldsymbol{\theta}})$ is obtained from (3.6) by changing $F(\boldsymbol{x}_k^T\hat{\boldsymbol{\alpha}})$ to 1 and $\boldsymbol{e}_k^*(\hat{\boldsymbol{\theta}})$ to $\boldsymbol{l}_k(\hat{\boldsymbol{\theta}})$. We also considered a naïve variance estimator $\vartheta_{DR,n}(\hat{\boldsymbol{\theta}})$ associated with $\hat{\boldsymbol{\theta}}$ which uses w_k in the place of d_k in $\vartheta_{DR}(\tilde{\boldsymbol{\theta}})$.

We generated $R = 10{,}000$ finite populations $\{y_k\}$, each of size $N = 393$, from $y_k \mid z_k \sim P(\lambda_k)$ and $z_k \sim B(1, p_k)$ with $\lambda_k = \exp(\theta_0 + z_k\theta_1)$ and $p_k = \exp(\delta_0 + \delta_1 x_k)/\{1+\exp(\delta_0 + \delta_1 x_k)\}$, where the fixed variable x_k is the number of beds in hospital k for the hospital population. The choice of $\delta_0 = 1$ and $\delta_1 = -.002$ leads to an average of about 60% for z_k. The group indicators z_k were generated for the first population and then fixed in the simulation of remaining populations, so that $\{z_k\}$ may be regarded as fixed explanatory variables. Our parameter of interest is $\boldsymbol{\theta} = (\theta_0, \theta_1)^T = (2,\ 1)^T$.

From each generated population, one simple random sample of size $n = 30$ was drawn. To implement post-stratification adjustment, we first grouped the population units into two classes with 271 units k having $x < 350$ in class 1 and 122 units k with $x \geq 350$ in class 2. Let u_{ck} be the class membership indicator for element k in class $c = 1,2$. We used GREG weight adjustment with $\boldsymbol{x}_k = (u_{1k}, u_{2k})^T$ and known $\boldsymbol{X} = (N_1, N_2)^T$, where $N_1 = 271$ and $N_2 = 122$.

We calculated the estimates $\hat{\boldsymbol{\theta}}$, $\tilde{\boldsymbol{\theta}}$ and the variance estimates $\vartheta_{DR}(\hat{\boldsymbol{\theta}})$, $\vartheta_{DR}(\tilde{\boldsymbol{\theta}})$ and $\vartheta_{DR,n}(\hat{\boldsymbol{\theta}})$ from each generated sample and their means $\hat{\boldsymbol{\theta}}_{\bullet}$, $\tilde{\boldsymbol{\theta}}_{\bullet}$, $\overline{\vartheta}_{DR}(\hat{\boldsymbol{\theta}})$, $\overline{\vartheta}_{DR}(\tilde{\boldsymbol{\theta}})$ and $\overline{\vartheta}_{DR,n}(\hat{\boldsymbol{\theta}})$, and the variances of $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$, denoted $V(\hat{\boldsymbol{\theta}})$ and $V(\tilde{\boldsymbol{\theta}})$. We have the following results:

(1) $V(\hat{\theta}_0) = .0139$ and $V(\hat{\theta}_1) = .0167$ compared to $V(\tilde{\theta}_0) = .0133$ and $V(\tilde{\theta}_1) = .0161$, suggesting that post-stratification is not effective for estimating model parameters when the model fits the data well; in fact, it lead to slight increase in variance. This result is in agreement with the observation made by Rao, Yung, and Hidiroglou (2002).

(2) $\overline{\vartheta}_{DR}(\hat{\theta}_0) = .0123$, $\overline{\vartheta}_{DR}(\hat{\theta}_1) = .0150$ compared to $V(\hat{\theta}_0) = .0139$ and $V(\hat{\theta}_1) = .0167$, suggesting that the variance estimator ϑ_{DR} tracks the corresponding total variance in the case of calibration. Similarly, $\overline{\vartheta}_{DR}(\tilde{\theta}_0) = .0122$ and $\overline{\vartheta}_{DR}(\tilde{\theta}_1) = .0148$ compared to $V(\tilde{\theta}_0) = .0133$ and $V(\tilde{\theta}_1) = .0161$, showing that ϑ_{DR} also tracks the total variance without calibration. Finally, the naive variance estimator of $\hat{\boldsymbol{\theta}}$ also tracks the total variance of $\hat{\boldsymbol{\theta}}$: $\overline{\vartheta}_{DR,n}(\hat{\theta}_0) = .0120$ and $\overline{\vartheta}_{DR,n}(\hat{\theta}_1) = .0145$ compared to $V(\hat{\theta}_0) = .0139$ and $V(\hat{\theta}_1) = .0167$.

4. Multiple Weight Adjustments

In the presence of missing responses, weighting adjustment is often used to compensate for complete nonresponse. Let r_k denotes the partial response indicator variable for the k^{th} element, i.e. $r_k = 0$ if there is complete nonresponse and $r_k = 1$ if there is partial response. A widely-used approach to adjust for complete nonresponse is to employ a new set of weights, $\tilde{\boldsymbol{w}}$, with k^{th} element equals to

$$\tilde{w}_k = d_k r_k F(\boldsymbol{x}_k^T\hat{\boldsymbol{\alpha}}), \tag{4.1}$$

where the vector parameter $\boldsymbol{\alpha}$, when predictor variables $\boldsymbol{x}_k = (x_{1k}, \ldots, x_{q_x k})^T$ are available for all sampled elements, is defined by solving a set of stochastic calibration constraints

$$\sum d_k \boldsymbol{x}_k p(\boldsymbol{x}_k^T\boldsymbol{\alpha}) = \sum d_k \boldsymbol{x}_k r_k, \tag{4.2}$$

where $p(\boldsymbol{x}_k^T\boldsymbol{\alpha}) = [lb + ub \times \exp(\boldsymbol{x}_k^T\boldsymbol{\alpha})]/[1+\exp(\boldsymbol{x}_k^T\boldsymbol{\alpha})]$, $F(\boldsymbol{x}_k^T\boldsymbol{\alpha}) = 1/p(\boldsymbol{x}_k^T\boldsymbol{\alpha})$, with lower and upper bounds, (lb, ub) generally set to (0,1). Note that (4.2) can be written as $\sum d_k \boldsymbol{x}_k (r_k - p(\boldsymbol{x}_k^T\boldsymbol{\alpha})) = \boldsymbol{0}$. Suppose an additional vector of calibration variables $\boldsymbol{t}_k = (t_{1k}, \ldots, t_{q_t k})^T$ with know totals $\boldsymbol{T} = (T_1, \ldots, T_{q_t})^T$ is available in addition to the vector $\boldsymbol{x}_k$. The vector $\boldsymbol{x}_k$ is assumed to be related to the response probability of element k, while the vector $\boldsymbol{t}_k$ is assumed to be related to the variables of interest. In this case the final weights are of the form

$$w_k = \tilde{w}_k \; G(\boldsymbol{t}_k^T \hat{\boldsymbol{\beta}}), \tag{4.3}$$

where the parameter $\boldsymbol{\beta}$ is determine by solving a set of constant calibration constraints

$$\sum \boldsymbol{t}_k \tilde{w}_k G(\boldsymbol{t}_k^T \boldsymbol{\beta}) = \boldsymbol{T}. \tag{4.4}$$

After adjustment for complete nonresponse and use of auxiliary information, the estimator $\hat{\boldsymbol{\theta}}$ of the model parameter $\boldsymbol{\theta}$ under the model $E_m(y_k) = \mu_k(\boldsymbol{\theta})$ is obtained as the solution to

$$\hat{\boldsymbol{l}}(\boldsymbol{\theta}) = \sum d_k r_k F(\boldsymbol{x}_k^T \hat{\boldsymbol{\alpha}}) G(\boldsymbol{t}_k^T \hat{\boldsymbol{\beta}}) \boldsymbol{l}_k(\boldsymbol{\theta}) = \boldsymbol{0}. \tag{4.5}$$

Let $\boldsymbol{d}_k = (d_{1k}, d_{2k}, \boldsymbol{d}_{3k}^T)^T$, where $d_{1k} = d_k$, $d_{2k} = d_k r_k$, and $\boldsymbol{d}_{3k} = d_k r_k \boldsymbol{l}_k(\boldsymbol{\theta})$. Then $\hat{\boldsymbol{\theta}}$ can be written as $\hat{\boldsymbol{\theta}} = f(\boldsymbol{A}_d)$, where $\boldsymbol{A}_d$ is a $(p+2)\times N$ matrix with k^{th} column $\boldsymbol{d}_k$. Following the implicit differentiation method of Demnati and Rao (2004), $\boldsymbol{z}_k = \partial f(\boldsymbol{A}_b)/\partial \boldsymbol{b}_k \,|_{A_b = A_d}$ is evaluated as:

$$\boldsymbol{z}_k^T = (\boldsymbol{z}_{1k}^T, \boldsymbol{z}_{2k}^T, \boldsymbol{z}_{3k}^T), \tag{4.6}$$

with

$$\boldsymbol{z}_{1k}^T = [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})]^{-1} \hat{\boldsymbol{B}}^T(\boldsymbol{e}^*, \hat{\boldsymbol{\alpha}}) \boldsymbol{x}_k p(\boldsymbol{x}_k^T \hat{\boldsymbol{\alpha}}),$$
$$\boldsymbol{z}_{2k}^T = -[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})]^{-1} \left(F(\boldsymbol{x}_k^T \hat{\boldsymbol{\alpha}}) G(\boldsymbol{t}_k^T \hat{\boldsymbol{\beta}}) \hat{\boldsymbol{B}}^T(\boldsymbol{l}, \hat{\boldsymbol{\beta}}) \boldsymbol{t}_k + \hat{\boldsymbol{B}}^T(\boldsymbol{e}^*, \hat{\boldsymbol{\alpha}}) \boldsymbol{x}_k \right)$$
$$\boldsymbol{z}_{3k}^T = [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})]^{-1} F(\boldsymbol{x}_k^T \hat{\boldsymbol{\alpha}}) G(\boldsymbol{t}_k^T \hat{\boldsymbol{\beta}}) \boldsymbol{I}_p,$$

where

$$\hat{\boldsymbol{B}}(\boldsymbol{l}, \boldsymbol{\beta}) = \left(\sum \tilde{w}_k g(\boldsymbol{t}_k^T \boldsymbol{\beta}) \boldsymbol{t}_k \boldsymbol{t}_k^T \right)^{-1} \sum \tilde{w}_k g(\boldsymbol{t}_k^T \boldsymbol{\beta}) \boldsymbol{t}_k \boldsymbol{l}_k^T(\hat{\boldsymbol{\theta}}),$$
$$\hat{\boldsymbol{B}}(\boldsymbol{e}^*, \boldsymbol{\alpha}) = [\hat{\boldsymbol{Q}}(\boldsymbol{\alpha})]^{-1} \sum d_k r_k [\partial p(\boldsymbol{x}_k^T \boldsymbol{\alpha})/\partial \boldsymbol{\alpha}] \boldsymbol{x}_k \boldsymbol{e}_k^{*T} / p^2(\boldsymbol{x}_k^T \boldsymbol{\alpha})$$
$$\hat{\boldsymbol{Q}}(\boldsymbol{\alpha}) = \sum d_k \boldsymbol{x}_k [\partial p(\boldsymbol{x}_k^T \boldsymbol{\alpha})/\partial \boldsymbol{\alpha}^T],$$

and

$$\boldsymbol{e}_k^* = G(\boldsymbol{t}_k^T \hat{\boldsymbol{\beta}}) [\boldsymbol{l}_k(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{B}}^T(\boldsymbol{l}, \hat{\boldsymbol{\beta}}) \boldsymbol{t}_k].$$

The DR linearization variance estimator of $\hat{\boldsymbol{\theta}}$ is given by

$$\vartheta_{DR}(\hat{\boldsymbol{\theta}}) = \sum\sum \boldsymbol{z}_k^T \operatorname{cov}(\boldsymbol{d}_k, \boldsymbol{d}_t) \boldsymbol{z}_t, \tag{4.7}$$

where

$$\begin{aligned} &\operatorname{cov}(\boldsymbol{d}_k, \boldsymbol{d}_t) = \\ &d_{kt} r_k r_t \hat{\xi}_k \hat{\xi}_t / \hat{\xi}_{kt} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \operatorname{cov}_m(\boldsymbol{l}_k(\hat{\boldsymbol{\theta}}), \boldsymbol{l}_t(\hat{\boldsymbol{\theta}})) \end{pmatrix} \\ &+ d_{kt} r_k r_t (\hat{\xi}_{kt} - \hat{\xi}_k \hat{\xi}_t) / \hat{\xi}_{kt} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & \boldsymbol{l}_t^T(\hat{\boldsymbol{\theta}}) \\ 0 & \boldsymbol{l}_k(\hat{\boldsymbol{\theta}}) & \boldsymbol{l}_k(\hat{\boldsymbol{\theta}}) \boldsymbol{l}_t^T(\hat{\boldsymbol{\theta}}) \end{pmatrix} \\ &+ d_{kt}(1-\omega_{kt})/\omega_{kt} \boldsymbol{v}_k \boldsymbol{v}_t^T, \end{aligned} \tag{4.8}$$

$\hat{\xi}_k = \hat{E}_r(r_k)$, $\hat{\xi}_{kt} = \hat{E}_r(r_k r_t)$, $\boldsymbol{v}_k = (1, r_k, r_k \boldsymbol{l}_k^T(\hat{\boldsymbol{\theta}}))^T$, and E_r is the response expectation. Substituting (4.8) in (4.7), we get

$$\begin{aligned} \vartheta_{DR}(\hat{\boldsymbol{\theta}}) &= \sum\sum d_{kt} r_k r_t [\hat{\xi}_k \hat{\xi}_t / \hat{\xi}_{kt}] \boldsymbol{z}_{k;m}^T \operatorname{cov}_m(\boldsymbol{l}_k(\hat{\boldsymbol{\theta}}), \boldsymbol{l}_t(\hat{\boldsymbol{\theta}})) \boldsymbol{z}_{t;m} \\ &+ \sum\sum d_{kt} r_k r_t [(\hat{\xi}_{kt} - \hat{\xi}_k \hat{\xi}_t)/\hat{\xi}_{kt}] \boldsymbol{z}_{k;r}^T \boldsymbol{z}_{t;r} \\ &+ \sum\sum d_{kt} [(1-\omega_{kt})/\omega_{kt}] \boldsymbol{z}_{k;s}^T \boldsymbol{z}_{t;s} \\ &\equiv \vartheta_m + \vartheta_r + \vartheta_s, \end{aligned} \tag{4.9}$$

where

$$\boldsymbol{z}_{k;m}^T = [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})]^{-1} F(\boldsymbol{x}_k^T \hat{\boldsymbol{\alpha}}) G(\boldsymbol{t}_k^T \hat{\boldsymbol{\beta}}) \boldsymbol{I}_p,$$
$$\boldsymbol{z}_{k;r}^T = [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})]^{-1} \left(F(\boldsymbol{x}_k^T \hat{\boldsymbol{\alpha}}) \boldsymbol{e}_k^* - \hat{\boldsymbol{B}}^T(\boldsymbol{e}^*, \hat{\boldsymbol{\alpha}}) \boldsymbol{x}_k \right), \text{ and}$$
$$\boldsymbol{z}_{k;s}^T = [\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})]^{-1} \left(r_k F(\boldsymbol{x}_k^T \hat{\boldsymbol{\alpha}}) \boldsymbol{e}_k^* - \hat{\boldsymbol{B}}^T(\boldsymbol{e}^*, \hat{\boldsymbol{\alpha}}) \boldsymbol{x}_k [r_k - p(\boldsymbol{x}_k^T \hat{\boldsymbol{\alpha}})] \right)$$

Note that the first component, ϑ_m, corresponds to the model, the second component, ϑ_r, corresponds to the response and the third component, ϑ_s, corresponds to the sampling design.

We generated $R = 1{,}000$ populations, each of size $N = 1{,}000$, using Hansen *et al*. (1983) model where $E_m(y\,|\,x) = 0.4 + 0.25x$ and $Var_m(y\,|\,x) = 0.0625x^{3/2}$ with both x and y having gamma distributions. The values of x were generated for the first population only and maintained fixed during the simulation. We used $\boldsymbol{x}_k^T \boldsymbol{\alpha} = 1 - 0.04\, x_k$, which leads to an average response rate of about 64% where $\boldsymbol{x}_k = (1, x_k)^T$. The estimating equation for element k is $l_k(\boldsymbol{\theta}) = \boldsymbol{c}_k(y_k - \boldsymbol{x}_k^T \boldsymbol{\theta})$ with $\boldsymbol{c}_k = (x_k^{-3/2}, x_k^{-1/2})^T$ and the vector parameter is $\boldsymbol{\theta} = (0.4, 0.25)^T \equiv (\theta_0, \theta_1)^T$. From each population one simple random sample of size $n = 100$ is drawn. To implement the adjustments we used in both adjustments the same vector of auxiliary variables $\boldsymbol{t}_k = \boldsymbol{x}_k$. For the complete nonresponse adjustment, we used (4.2) with $(lb, ub) = (0,1)$; and for the second weight adjustment, we used $\boldsymbol{T} = (N, X)^T$ in combination with $G(\boldsymbol{t}_k^T \boldsymbol{\beta}) = \dfrac{l(u-1) + u(1-l) \times \exp(A\boldsymbol{t}_k^T \boldsymbol{\beta})}{(u-1) + (1-l)\exp(A\boldsymbol{t}_k^T \boldsymbol{\beta})}$, where $A = (u-l)/[(u-1)(1-l)]$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, $l = 0.6$ and $u = 1.4$. The estimator of model parameter is given by $\hat{\boldsymbol{\theta}} = [\sum w_k \boldsymbol{c}_k \boldsymbol{x}_k^T]^{-1} \sum w_k \boldsymbol{c}_k y_k$. Only 720 cases converged. The nonconvergence is mainly due to nonresponse adjustment. For the 720 cases, the bias in the estimation of $\boldsymbol{\theta}$ is negligible: $\theta_0 = 0.4$ vs. $\bar{\hat{\theta}}_0 = 0.40248$ and $\theta_1 = 0.25$ vs. $\bar{\hat{\theta}}_1 = 0.25069$, where $\bar{\hat{\theta}}_0$ and $\bar{\hat{\theta}}_1$ denote the average values of $\hat{\theta}_0$ and $\hat{\theta}_1$. We have the following results on the average values $\bar{\vartheta}_{DR}$, $\bar{\vartheta}_m$, $\bar{\vartheta}_r$, and $\bar{\vartheta}_s$ compared to simulated $V(\hat{\boldsymbol{\theta}})$:

$V(\hat{\theta}_0) = .015$, $\bar{\vartheta}_{DR}(\hat{\theta}_0) = .021$, $\bar{\vartheta}_s(\hat{\theta}_0) = .020$, $\bar{\vartheta}_r(\hat{\theta}_0) = .0004$, $\bar{\vartheta}_m(\hat{\theta}_0) = .0009$,

$V(\hat{\theta}_1) = .0007$, $\bar{\vartheta}_{DR}(\hat{\theta}_1) = .0009$, $\bar{\vartheta}_s(\hat{\theta}_1) = .0009$, $\bar{\vartheta}_r(\hat{\theta}_1) = .00002$, $\bar{\vartheta}_m(\hat{\theta}_1) = .00003$

$V(\hat{\theta}_0, \hat{\theta}_1) = -.0023$, $\bar{\vartheta}_{DR}(\hat{\theta}_0, \hat{\theta}_1) = -.0031$, $\bar{\vartheta}_s(\hat{\theta}_0, \hat{\theta}_1) = -.0029$, $\bar{\vartheta}_r(\hat{\theta}_0, \hat{\theta}_1) = -.00008$, $\bar{\vartheta}_m(\hat{\theta}_0, \hat{\theta}_1) = -.0001$. The percent coverage of 95% normal theory confidence intervals associated with ϑ_{DR} are 94.4 for θ_0 and 96.3 for θ_1.

References

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, pp. 736-740.

Demnati, A. and Rao, J. N. K. (2003), “Linearization Variance Estimators for Longitudinal Survey Data”, *Federal Committee on Statistical Methodology Research Conference*, pp. 138-143.

Demnati, A. and Rao, J. N. K. (2004), “Linearization Variance Estimators for Survey Data (with discussion)”, *Survey Methodology*, 30, pp. 17-34.

Demnati, A. and Rao, J. N. K. (2005), “Linearization Variance Estimators for Model Parameters from Complex Survey Data”, *Proceedings of Statistics Canada Symposium*.

Hansen, M. H., Madow, W. G., and Tepping, B. J. (1983), “An Evaluation of Model-Dependent and Probability Sampling Inferences in Sample Surveys”. *Journal of the American Statistical Association,* 78, 776-793.

Rao, J.N.K., Yung, W. and Hidiroglou, M. (2002), “Estimating Equations for the Analysis of Survey Data using Poststratification Information”, *Sankyhā: The Indian Journal of Statistics*, 64, 1-15.

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, pp. 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, pp. 527-537.

Valliant, R. (1993), “Postsratification and Conditional Variance Estimation”, *Journal of the American Statistical Association*, 88, pp. 89-96.

Valliant R., Dorfman, A. H. and Royall, R. M. (2000) “*Finite Population Sampling and Inference: A Prediction Approach*”, Wiley.

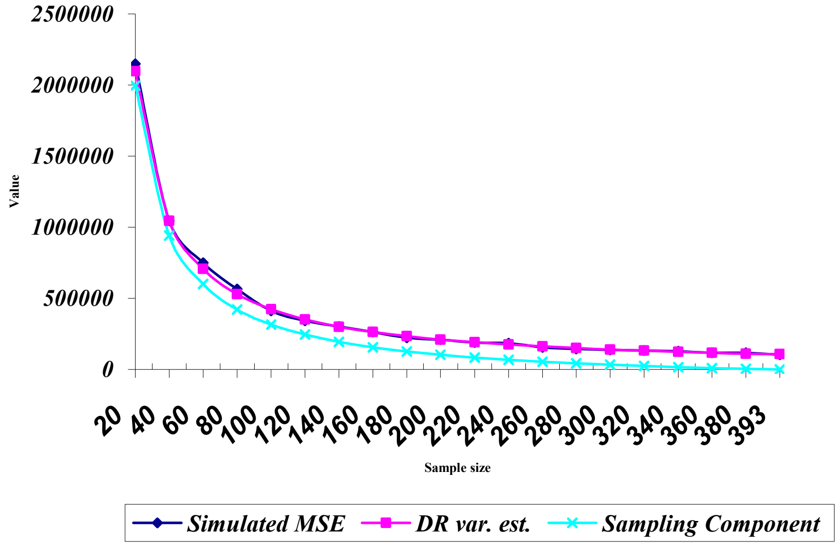

Figure 1: Averages of variance estimates for selected sample sizes compared to estimated MSE of the ratio estimator. $\overline{\vartheta}_{DR}$=*DR var. est.*, $\overline{\vartheta}_{s}$ = *Sampling component*

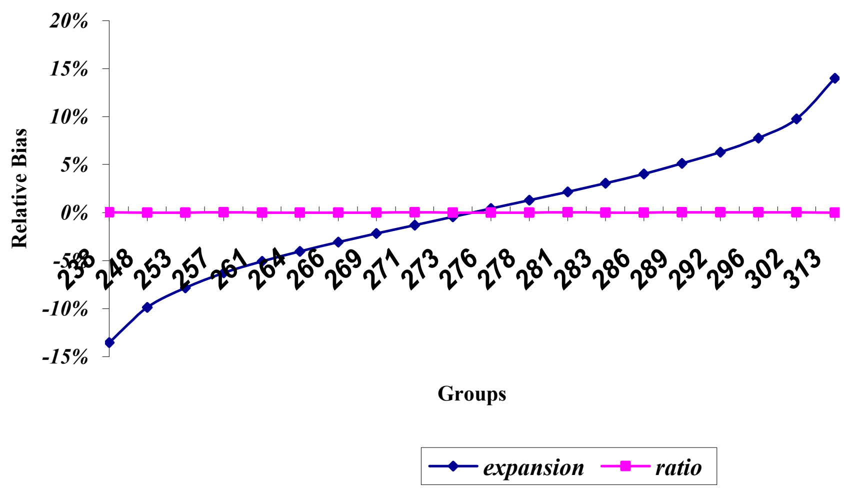

Figure 2: Conditional relative bias of the expansion and ratio estimators

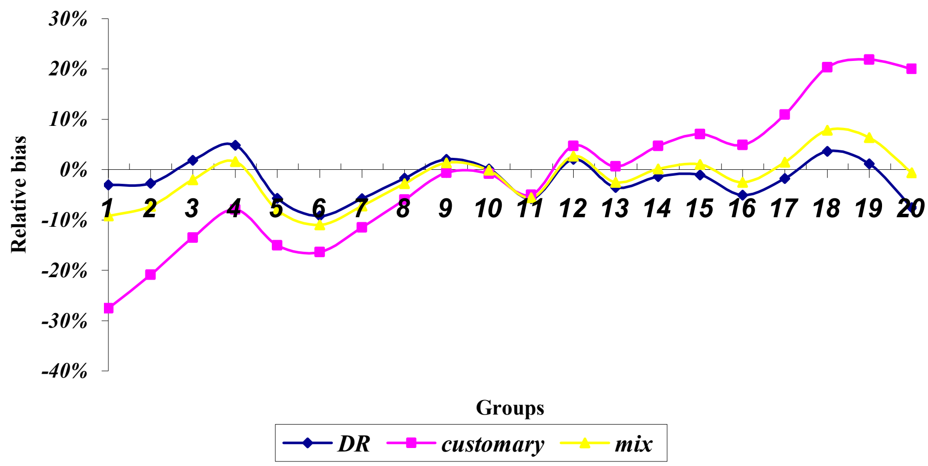

Figure 3: Conditional relative bias of variance estimators ϑ_{DR}, ϑ_{cus} and ϑ_{mix}

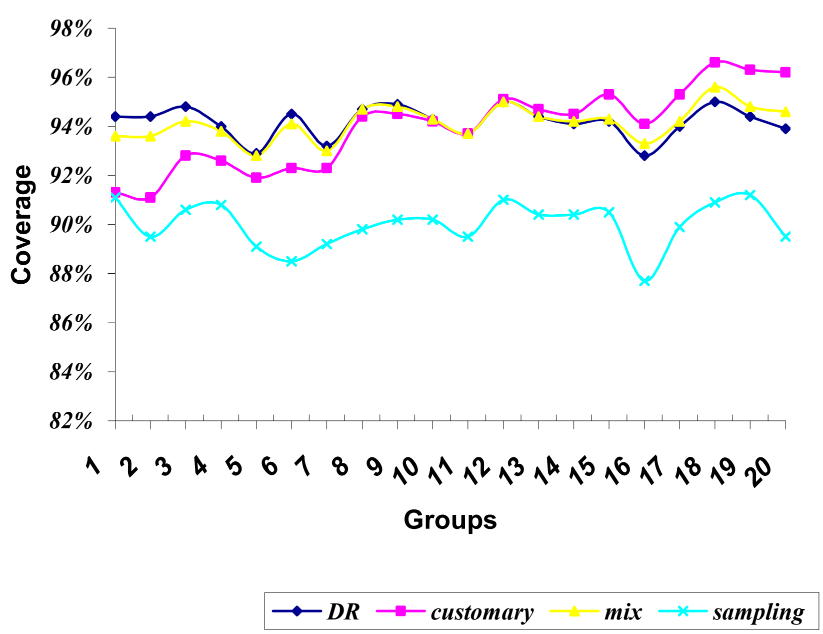

Figure 4: Conditional coverage rates of normal theory confidence intervals based on ϑ_{DR}, ϑ_{cus}, ϑ_{mix} and ϑ_s for nominal level of 95%