Statistica Sinica **24** (2014), 375-394
doi:http://dx.doi.org/10.5705/ss.2012.005

DOUBLY ROBUST INFERENCE WITH MISSING DATA IN SURVEY SAMPLING

Jae Kwang Kim and David Haziza

Iowa State University and Université de Montréal

Abstract: Statistical inference with missing data requires assumptions about the population or about the response probability. Doubly robust (DR) estimators use both relationships to estimate the parameters of interest, so that they are consistent even when one of the models is misspecified. In this paper, we propose a method of computing propensity scores that leads to DR estimation. In addition, we discuss DR variance estimation so that the resulting inference is doubly robust. Some asymptotic properties are discussed. Results from two limited simulation studies are also presented.

Key words and phrases: Calibration, double protection, nonresponse, variance estimation.

1. Introduction

Missing data occurs in surveys because some of the sampled units refuse to respond to the survey or because of the inability to contact them. Dropout or noncompliance in clinical trials may also lead to missing responses for some subjects. It is well known that unadjusted estimators may be heavily biased if the respondents differ from the nonrespondents systematically with respect to the study variables. It is thus desirable to develop estimation procedures exhibiting low biases.

To adjust for the bias associated with missing data, two modeling approaches are often used: the response probability (RP) model approach that requires the specification of a response model describing the unknown nonresponse mechanism and the outcome regression (OR) model approach that requires the specification of the model describing the distribution of the study variable. In survey sampling the RP model approach is also called the nonresponse model approach, whereas the OR model approach is called the prediction model approach or the imputation model approach. An estimator is said to be doubly robust (DR) if it remains asymptotically unbiased and consistent if either model (nonresponse or outcome regression) is true. DR procedures offer some protection against misspecification of one model or the other. This is clearly an attractive property and is closely related with the philosophy of model-assisted estimation in survey sampling (Särndal, Swensson and Wretman (1992); Firth and Bennett (1998); Fuller (2009)).

In recent years, DR estimation procedures have attracted a lot of attention in mainstream statistics; e.g., Robins, Rotnitzky, and Zhao (1994), Scharfstein, Rotnitzky, and Robins (1999), Tan (2006), Bang and Robins (2005), Kang and Schafer (2008), Cao, Tsiatis, and Davidian (2009), among others. In the survey sampling context, DR estimation has been studied in Kott (1994), Kott (2006), Kott and Chang (2010), Kim and Park (2006), and Haziza and Rao (2006), among others. Kott (2006) discussed the doubly robustness of the variance estimator proposed by Folsom and Singh (2000) in the context of calibration for unit nonresponse in survey sampling; see also Kott and Chang (2010). In the context of imputation for missing data, DR variance estimation has been discussed in Haziza and Rao (2006) and Kim and Park (2006) when the overall sampling fraction is negligible. Haziza and Rao (2006) considered Taylor linearization procedures, whereas replication variance estimation was studied in Kim and Park (2006). However, Haziza and Rao (2006) and Kim and Park (2006) did not address the double robustness of the variance estimators for large sampling fractions.

We consider DR inference in the sense that the inference based on point estimator and variance estimator is justified if either one of the two models, nonresponse model or outcome regression model, holds. The proposed doubly robust estimator is quite efficient and provides a variance estimator that can be easily implemented using software designed for complete data variance estimation. In finite population sampling, the proposed variance estimator is slightly modified.

In Section 2, the basic setup is introduced. The proposed DR estimator and its variance estimator are discussed in Section 3. In Section 4, the proposed method is applied to the survey sampling context and the proposed variance estimator is modified to account for the finite population. Results from two simulation studies are presented in Section 5 to compare the performance of the proposed estimator with those from existing methods. Concluding remarks are made in Section 6.

2. Basic Setup

Suppose we have n independent realizations of a random variable Y, $y_1, \ldots,$ y_n, from some distribution, and that we are interested in estimating $\theta = E(Y)$. In the absence of nonresponse to the study variable y, the parameter θ is consistently estimated by the sample mean

$$\hat{\theta}_n = \sum_{i=1}^{n} w_i y_i, \tag{2.1}$$

where $w_i = 1/n$. In Section 4, we use a different set of weights w_i as we treat the problem of DR inference in the survey sampling context.

In addition to the study variable y, suppose a vector of auxiliary variables, $\mathbf{x}$, is available in the sample. Let δ_i be a response indicator attached to unit i such that $\delta_i = 1$ if y_i is observed and $\delta_i = 0$, otherwise. Instead of observing $(\mathbf{x}_i, y_i)$ for the whole sample, we observe $(\mathbf{x}_i, y_i)$ for $\delta_i = 1$ and observe only $\mathbf{x}_i$ for $\delta_i = 0$. We assume that the response mechanism is missing at random (MAR) in the sense of Rubin (1976).

A natural approach for estimating θ consists of first postulating a model for the conditional distribution of y_i given $\mathbf{x}_i$. In particular, if we are only interested in the mean of the y-values, we consider the following model

$$E\left(y_i \mid \mathbf{x}_i, \delta_i = 0\right) = m\left(\mathbf{x}_i; \boldsymbol{\beta}_0\right), \tag{2.2}$$

where $m\left(\mathbf{x}_i, \boldsymbol{\beta}\right)$ is a continuous differentiable function of $\boldsymbol{\beta}$. The model (2.2) is called the OR model. Under MAR, (2.2) implies that $E\left(y_i \mid \mathbf{x}_i\right) = m\left(\mathbf{x}_i; \boldsymbol{\beta}_0\right)$.

A natural estimator of θ is the (deterministically) imputed estimator

$$\hat{\theta}_p = \sum_{i=1}^{n} w_i \left\{ \delta_i y_i + \left(1 - \delta_i\right) m(\mathbf{x}_i, \hat{\boldsymbol{\beta}}) \right\}, \tag{2.3}$$

where $\hat{\boldsymbol{\beta}}$ is a consistent estimator of the true parameter $\boldsymbol{\beta}_0$.

Since

$$\hat{\theta}_p - \hat{\theta}_n = -\sum_{i=1}^{n} w_i \left(1 - \delta_i\right) \left\{ y_i - m(\mathbf{x}_i, \hat{\boldsymbol{\beta}}), \right\}$$

we have

$$\begin{aligned} &E\left\{ \hat{\theta}_p - \hat{\theta}_n \mid \delta_1, \ldots, \delta_n, \mathbf{x}_1, \ldots, \mathbf{x}_n \right\} \\ &= -\sum_{i=1}^{n} w_i \left(1 - \delta_i\right) \left\{ E\left(y_i \mid \mathbf{x}_i, \delta_i = 0\right) - m(\mathbf{x}_i, \hat{\boldsymbol{\beta}}) \right\}, \end{aligned}$$

where $E(\cdot \mid \delta_1, \ldots, \delta_n, \mathbf{x}_1, \ldots, \mathbf{x}_n)$ denotes the conditional expectation with respect to the OR model. Thus, the validity of the imputed estimator (2.3) follows if (2.2) is true and $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}_0$.

Now, suppose that the probability of response to the study variable y, $p_i = \Pr\left(\delta_i = 1 \mid \mathbf{x}_i\right)$, follows a parametric model

$$p_i = p_i(\boldsymbol{\phi}_0) = \frac{\exp\left(\boldsymbol{\phi}_0' \mathbf{x}_i\right)}{1 + \exp\left(\boldsymbol{\phi}_0' \mathbf{x}_i\right)} \tag{2.4}$$

for some $\boldsymbol{\phi}_0$. The model (2.4) is called the RP model. We assume that the intercept term is included in (2.4). In the classical two-phase sampling setup, where the second-phase sample corresponds to the set of respondents, the second-phase conditional inclusion probability p_i is known and the two-phase regression estimator

$$\hat{\theta}_{tp} = \sum_{i=1}^{n} w_i \Big[m(\mathbf{x}_i; \hat{\boldsymbol{\beta}}) + \frac{\delta_i}{p_i} \Big\{ y_i - m(\mathbf{x}_i; \hat{\boldsymbol{\beta}}) \Big\} \Big]$$
$$= \hat{\theta}_n + \sum_{i=1}^{n} w_i \Big(\frac{\delta_i}{p_i} - 1 \Big) \big\{ y_i - m\big(\mathbf{x}_i; \hat{\boldsymbol{\beta}}\big) \big\} \quad (2.5)$$

is approximately unbiased for θ under the nonresponse model (Cochran (1977)) regardless of whether or not (2.2) holds. When the RP model is not correct, the estimator is still approximately unbiased if (2.2) and the MAR condition hold and $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}_0$. Thus, $\hat{\theta}_{tp}$ is doubly robust in the sense that it remains valid if either one of the two models holds.

When the response probability is estimated, rather than known, we consider a class of estimators of the form

$$\hat{\theta}_{DR}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}) = \hat{\theta}_n + \sum_{i=1}^{n} w_i \Big\{ \frac{\delta_i}{p_i(\hat{\boldsymbol{\phi}})} - 1 \Big\} \Big\{ y_i - m\left(\mathbf{x}_i; \hat{\boldsymbol{\beta}} \right) \Big\}, \quad (2.6)$$

indexed by $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$, where $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}_0$ under the assumed OR model and $\hat{\boldsymbol{\phi}}$ is consistent for $\boldsymbol{\phi}_0$ under the assumed RP model. As noted by Scharfstein, Rotnitzky, and Robins (1999), the double robustness property also follows if p_i is replaced by $\hat{p}_i = p_i(\hat{\boldsymbol{\phi}})$ using a consistent estimator $\hat{\boldsymbol{\phi}}$ for $\boldsymbol{\phi}_0$. Note that the doubly robust estimator, $\hat{\theta}_{DR}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$, in (2.6) is a class of estimators and different choices of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$ lead to different doubly robust estimators. Scharfstein, Rotnitzky, and Robins (1999) and Haziza and Rao (2006) used $\hat{\boldsymbol{\phi}}$ estimated by maximum likelihood and $\hat{\boldsymbol{\beta}}$ estimated by ordinary or iteratively reweighted least squares. Recently, Cao, Tsiatis, and Davidian (2009) proposed a doubly robust estimator using the optimal score equation based on influence function theory. However, the proposed variance estimator of Cao, Tsiatis, and Davidian (2009) is not necessarily doubly robust.

We propose a DR estimator of the form (2.6) using a different choice of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$ which leads to a simplified DR variance estimator. Thus, the proposed point and variance estimation procedure leads to DR inference.

3. Main Results

Let

$$S(\phi) \equiv \sum_{i=1}^{n} w_i \Big\{ \frac{\delta_i}{p_i(\boldsymbol{\phi})} - 1 \Big\} \mathbf{h}_i(\phi) = 0 \quad (3.1)$$

be the (weighted) score equation for $\boldsymbol{\phi}_0$, where $\mathbf{h}_i(\phi) = \{\partial p_i(\boldsymbol{\phi})/\partial \boldsymbol{\phi}\} / \{1 - p_i(\boldsymbol{\phi})\}$. Given the choice of $\hat{p}_i = p_i(\hat{\phi}_{MLE})$ where $\hat{\boldsymbol{\phi}}_{MLE}$ satisfies (3.1), Cao, Tsiatis, and Davidian (2009) considered so-called the optimal DR estimator among the class of the estimators of the form

$$\hat{\theta}_{DR}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} w_i \Big\{ m\left(\mathbf{x}_i; \hat{\boldsymbol{\beta}}\right) + \frac{\delta_i}{\hat{p}_i}(y_i - m(\mathbf{x}_i; \hat{\boldsymbol{\beta}})) \Big\}. \tag{3.2}$$

Rubin and van der Laan (2008) considered the $\hat{\boldsymbol{\beta}}$ that minimizes

$$\sum_{i=1}^{n} w_i^2 \frac{\delta_i}{\hat{p}_i} \Big(\frac{1}{\hat{p}_i} - 1 \Big) \left\{ y_i - m\left(\mathbf{x}_i; \boldsymbol{\beta}\right) \right\}^2 ,$$

which is essentially the conditional variance ignoring the effect of estimating $\boldsymbol{\phi}_0$ in the estimated response probability $\hat{p}_i = p_i(\hat{\boldsymbol{\phi}})$. To correctly account for the effect of estimating $\boldsymbol{\phi}_0$, Cao, Tsiatis, and Davidian (2009) proposed minimizing

$$\sum_{i=1}^{n} w_i^2 \frac{\delta_i}{\hat{p}_i} \Big(\frac{1}{\hat{p}_i} - 1 \Big) \left\{ y_i - m\left(\mathbf{x}_i; \boldsymbol{\beta}\right) - \mathbf{c}'\mathbf{h}_i(\hat{\boldsymbol{\phi}}_{MLE}) \right\}^2$$

with respect to $(\boldsymbol{\beta}, \mathbf{c})$. Note that in this case there is no guarantee that the resulting estimator is optimal under the OR model. In fact, the proposed estimator of Cao, Tsiatis, and Davidian (2009) is sub-optimal because they first estimate $\hat{\boldsymbol{\phi}}$ by $\hat{\boldsymbol{\phi}}_{MLE}$ obtained from maximum likelihood and then seek for the optimal estimator in the class of estimators $\hat{\theta}^*_{DR}(\hat{\boldsymbol{\beta}}) = \hat{\theta}_{DR}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}_{MLE})$ as a function of $\hat{\boldsymbol{\beta}}$. As discussed in Kim and Kim (2007) and Kim and Riddles (2012), the choice of $\hat{\boldsymbol{\phi}}_{MLE}$ does not necessarily lead to the optimal propensity score estimators. For example, according to Kim and Riddles (2012), when the OR model is $m(\mathbf{x}_i; \boldsymbol{\beta}) = \mathbf{x}_i'\boldsymbol{\beta}$, the optimal choice of $\hat{\boldsymbol{\phi}}$ can be obtained by solving

$$\sum_{i=1}^{n} w_i \frac{\delta_i}{p_i\left(\boldsymbol{\phi}\right)} \mathbf{x}_i = \sum_{i=1}^{n} w_i \mathbf{x}_i, \tag{3.3}$$

which is different from the score equation for the MLE of $\boldsymbol{\phi}_0$. Thus, we expect that the efficiency of the sub-optimal estimator of Cao, Tsiatis, and Davidian (2009) can be improved for a suitable choice of $\hat{\boldsymbol{\phi}}$.

We propose a DR estimator $\hat{\theta}_p$ of the form (2.3) using $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$, where $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$ is obtained by solving

$$\sum_{i=1}^{n} w_i \delta_i \Big\{ \frac{1}{p_i\left(\boldsymbol{\phi}\right)} - 1 \Big\} \left\{ y_i - m\left(\mathbf{x}_i; \boldsymbol{\beta}\right) \right\} \mathbf{x}_i = \mathbf{0}, \tag{3.4}$$

$$\sum_{i=1}^{n} w_i \Big\{ \frac{\delta_i}{p_i\left(\boldsymbol{\phi}\right)} - 1 \Big\} \dot{m}\left(\mathbf{x}_i; \boldsymbol{\beta}\right) = \mathbf{0}, \tag{3.5}$$

simultaneously, where $\dot{m}\left(\mathbf{x}_i; \boldsymbol{\beta}\right) = \partial m(\mathbf{x}_i; \boldsymbol{\beta}) / \partial \boldsymbol{\beta}$. Because an intercept term is included in $\mathbf{x}$, (3.4) implies that

$$\sum_{i=1}^{n} w_i \delta_i \frac{1}{p_i(\hat{\boldsymbol{\phi}})} \big\{ y_i - m(\mathbf{x}_i; \hat{\boldsymbol{\beta}}) \big\} = \sum_{i=1}^{n} w_i \delta_i \big\{ y_i - m(\mathbf{x}_i; \hat{\boldsymbol{\beta}}) \big\}.$$

Thus, by (3.4), the imputed estimator (2.3) can be expressed as a doubly robust estimator of the form (2.6).

Condition (3.4) has been used in Scharfstein, Rotnitzky, and Robins (1999) and Haziza and Rao (2006). Condition (3.5) is a calibration condition in the sense that the propensity score adjusted estimator applied to $\dot{m}\left(\mathbf{x}_i;\boldsymbol{\beta}\right)$ leads to the complete sample estimator. For example, consider the linear OR model for which $m(\mathbf{x}_i;\boldsymbol{\beta}) = \mathbf{x}_i'\boldsymbol{\beta}$. Then, (3.5) is equivalent to (3.3). Condition (3.3) was considered by Folsom (1991), Iannacchione, Milne, and Folsom (1991), and Chang and Kott (2008) in the context of unit nonresponse in survey sampling. From (3.3), it follows that estimates corresponding to the x-variables do not suffer from nonresponse error. Condition (3.4) means that $\hat{\boldsymbol{\beta}}$ is computed as

$$\hat{\boldsymbol{\beta}} = \Big\{\sum_{i=1}^{n}\delta_i(\hat{p}_i^{-1}-1)\mathbf{x}_i\mathbf{x}_i'\Big\}^{-1}\sum_{i=1}^{n}\delta_i(\hat{p}_i^{-1}-1)\mathbf{x}_i y_i.$$

Writing $y_i = \mathbf{x}_i'\boldsymbol{\beta}_0 + e_i$, the imputed estimator $\hat{\theta}_p$ can be written as

$$\hat{\theta}_p = \hat{\theta}_n + \sum_{i=1}^{n} w_i\Big\{\frac{\delta_i}{p_i(\hat{\boldsymbol{\phi}})}-1\Big\}\mathbf{x}_i'\left(\boldsymbol{\beta}_0-\hat{\boldsymbol{\beta}}\right) + \sum_{i=1}^{n} w_i\Big\{\frac{\delta_i}{p_i(\hat{\boldsymbol{\phi}})}-1\Big\}e_i.$$

Note that the second term here is zero if (3.5) holds. Thus, under (3.5),

$$\hat{\theta}_p = \hat{\theta}_n + \sum_{i=1}^{n} w_i\Big\{\frac{\delta_i}{p_i(\hat{\boldsymbol{\phi}})}-1\Big\}e_i$$

and the variability associated with $\hat{\boldsymbol{\beta}}$ can be safely ignored. Furthermore, using the fact $\partial p_i^{-1}\left(\boldsymbol{\phi}\right)/\partial\boldsymbol{\phi} = -\left\{p_i^{-1}\left(\boldsymbol{\phi}\right)-1\right\}\mathbf{x}_i$ under (2.4), we can apply a Taylor expansion to get

$$\hat{\theta}_p = \hat{\theta}_n + \sum_{i=1}^{n} w_i\Big\{\frac{\delta_i}{p_i(\boldsymbol{\phi}^*)}-1\Big\}e_i - \sum_{i=1}^{n} w_i\delta_i\Big\{\frac{1}{p_i(\boldsymbol{\phi}^*)}-1\Big\}e_i\mathbf{x}_i\left(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}^*\right) + O_p\left(n^{-1}\right), \tag{3.6}$$

where $\boldsymbol{\phi}^*$ is the probability limit of $\hat{\boldsymbol{\phi}}$. Using (3.4), it can be shown that

$$\sum_{i=1}^{n} w_i\delta_i\Big\{\frac{1}{p_i(\boldsymbol{\phi}^*)}-1\Big\}e_i\mathbf{x}_i = o_p\left(1\right)$$

and (3.6) reduces to

$$\hat{\theta}_p = \hat{\theta}_n + \sum_{i=1}^{n} w_i\Big\{\frac{\delta_i}{p_i(\boldsymbol{\phi}^*)}-1\Big\}e_i + o_p\left(n^{-1/2}\right). \tag{3.7}$$

Thus, the variability associated with $\hat{\boldsymbol{\phi}}$ can also be safely ignored.

We need some regularity conditions. The following theorem extends the above results to the general form of $E(y_i \mid \mathbf{x}_i) = m(\mathbf{x}_i; \boldsymbol{\beta}_0)$.

Assume the following regularity conditions:

(C.1) There is a fixed constant K_B such that $p_i^{-1} < K_B$ for all $i = 1, 2, \ldots, n$.

(C.2) The response probability function $p_i(\phi)$ is differentiable with continuous first order partial derivatives for all ϕ.

(C.3) The solution $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$ to (3.4) and (3.5) is uniquely determined and satisfies $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}) = (\boldsymbol{\beta}^*, \boldsymbol{\phi}^*) + o_p(1)$ for some $(\boldsymbol{\beta}^*, \boldsymbol{\phi}^*)$.

(C.4) The mean function $m(\mathbf{x}_i; \boldsymbol{\beta})$ is twice differentiable with continuous second-order partial derivatives for all $\boldsymbol{\beta}$.

(C.5) $W(\boldsymbol{\beta}) = (X, Y, m(\mathbf{x}; \boldsymbol{\beta}), \dot{m}(\mathbf{x}; \boldsymbol{\beta}))$ has finite fourth moment for all $\boldsymbol{\beta}$.

Theorem 1. *Under* (C.1)−(C.5)*, we have*

$$\sqrt{n}\left(\hat{\theta}_p - \tilde{\theta}_p\right) = o_p(1), \tag{3.8}$$

where

$$\tilde{\theta}_p = \sum_{i=1}^{n} w_i \Big[m(\mathbf{x}_i; \boldsymbol{\beta}^*) + \frac{\delta_i}{p_i(\boldsymbol{\phi}^*)} \{ y_i - m(\mathbf{x}_i; \boldsymbol{\beta}^*) \} \Big] \tag{3.9}$$

and $(\boldsymbol{\beta}^*, \boldsymbol{\phi}^*)$ *is the probability limit of* $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$.

Proof. Write the DR estimator as $\hat{\theta}_p = \hat{\theta}_p(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$, where $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$ is the solution to (3.4) and (3.5). Now, if

$$U(\boldsymbol{\beta}, \boldsymbol{\phi}) = \sum_{i=1}^{n} w_i \Big(\frac{\delta_i}{p_i(\boldsymbol{\phi})} - 1 \Big) \{ y_i - m(\mathbf{x}_i; \boldsymbol{\beta}) \},$$

we can write

$$\hat{\theta}_p(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}) = \hat{\theta}_n + U(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}). \tag{3.10}$$

Note that $U(\boldsymbol{\beta}, \boldsymbol{\phi})$ satisfies

$$\frac{\partial}{\partial \phi} U(\boldsymbol{\beta}, \boldsymbol{\phi}) = -\sum_{i=1}^{n} w_i \delta_i \Big\{ \frac{1 - p_i(\boldsymbol{\phi})}{p_i(\boldsymbol{\phi})} \Big\} \{ y_i - m(\mathbf{x}_i; \boldsymbol{\beta}) \} \mathbf{x}_i,$$

$$\frac{\partial}{\partial \boldsymbol{\beta}} U(\boldsymbol{\beta}, \boldsymbol{\phi}) = -\sum_{i=1}^{n} w_i \Big\{ \frac{\delta_i}{p_i(\boldsymbol{\phi})} - 1 \Big\} \dot{m}(\mathbf{x}_i; \boldsymbol{\beta}).$$

Thus, conditions (3.4) and (3.5), are equivalent to

$$\frac{\partial}{\partial (\boldsymbol{\beta}, \boldsymbol{\phi})} U(\boldsymbol{\beta}, \boldsymbol{\phi}) = \mathbf{0}. \tag{3.11}$$

Because of the existence of the second moment of the partial derivatives in (3.11), standard arguments for the asymptotic normality of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$ can be used to show that

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}) - (\boldsymbol{\beta}^*, \boldsymbol{\phi}^*) = O_p\left(n^{-1/2}\right). \tag{3.12}$$

Because $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$ satisfies (3.11), its probability limit $(\boldsymbol{\beta}^*, \boldsymbol{\phi}^*)$ satisfies

$$E\Big\{\frac{\partial}{\partial(\boldsymbol{\beta}, \boldsymbol{\phi})} U\left(\boldsymbol{\beta}, \boldsymbol{\phi}\right) \mid \boldsymbol{\beta} = \boldsymbol{\beta}^*, \boldsymbol{\phi} = \boldsymbol{\phi}^*\Big\} = \mathbf{0}. \tag{3.13}$$

Condition (3.13) that implies that the contribution due to estimating the parameters $(\boldsymbol{\beta}, \boldsymbol{\phi})$ is negligible in the asymptotic distribution of $U\left(\boldsymbol{\beta}, \boldsymbol{\phi}\right)$, is often called Randles (1982) condition. From (3.12) and (3.13), we obtain

$$U(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}) = U(\boldsymbol{\beta}^*, \boldsymbol{\phi}^*) + o_p(n^{-1/2}). \tag{3.14}$$

Therefore, combining (3.10) and (3.14), we have (3.8).

The probability statement in (3.8) is made in the doubly robust sense that the convergence in probability holds if one of the two models is true. If the reference distribution in (3.8) is with respect to (2.2), then $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$. If the reference distribution in (3.8) is with respect to (2.4), then $\boldsymbol{\phi}^* = \boldsymbol{\phi}_0$. When the two models are true, then $(\boldsymbol{\beta}^*, \boldsymbol{\phi}^*) = (\boldsymbol{\beta}_0, \boldsymbol{\phi}_0)$ and the variance of $\tilde{\theta}_p$ is

$$V\left(\tilde{\theta}_p\right) = V\left(\hat{\theta}_n\right) + E\Big\{\sum_{i=1}^n w_i^2 \{p_i(\boldsymbol{\phi}_0)^{-1} - 1\} e_i^2\Big\}, \tag{3.15}$$

where $e_i = y_i - m(\mathbf{x}_i; \boldsymbol{\beta}_0)$. Under simple random sampling, (3.15) is equal to the semiparametric lower bound of the asymptotic variance and, as a result, $\hat{\theta}_p$ is locally efficient (Robins, Rotnitzky, and Zhao (1994)).

Taking

$$\eta_i\left(\boldsymbol{\beta}, \boldsymbol{\phi}\right) = m\left(\mathbf{x}_i; \boldsymbol{\beta}\right) + \frac{\delta_i}{p_i\left(\boldsymbol{\phi}\right)} \left\{y_i - m\left(\mathbf{x}_i; \boldsymbol{\beta}\right)\right\}, \tag{3.16}$$

(3.8) means that

$$\sum_{i=1}^n w_i \eta_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}) = \sum_{i=1}^n w_i \eta_i\left(\boldsymbol{\beta}^*, \boldsymbol{\phi}^*\right) + o_p\left(n^{-1/2}\right).$$

Thus, if $(\mathbf{x}_i, y_i, \delta_i)$ are i.i.d., the $\eta_i(\boldsymbol{\beta}^*, \boldsymbol{\phi}^*)$ are i.i.d., even though $\eta_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$ are not necessarily i.i.d.. Because $\eta_i\left(\boldsymbol{\beta}^*, \boldsymbol{\phi}^*\right)$ are i.i.d., we can apply the Central Limit Theorem and the Slutsky Theorem to get

$$\sqrt{n}\left(\hat{\theta}_p - \theta\right) \xrightarrow{L} N\left(0, \sigma^2\right), \tag{3.17}$$

where $\xrightarrow{L}$ denotes the convergence in distribution and $\sigma^2 = Var\{\eta_i(\boldsymbol{\beta}^*, \boldsymbol{\phi}^*)\}$. Furthermore, since $\eta_i(\boldsymbol{\beta}^*, \boldsymbol{\phi}^*)$ are i.i.d with bounded fourth moments, we can apply the standard complete sample method to estimate the variance of $\tilde{\theta}_p = \sum_{i=1}^n w_i \eta_i (\boldsymbol{\beta}^*, \boldsymbol{\phi}^*)$. Then

$$\hat{V}(\boldsymbol{\beta}^*, \boldsymbol{\phi}^*) = \frac{1}{n}\frac{1}{n-1}\sum_{i=1}^n (\eta_i - \bar{\eta}_n)^2, \tag{3.18}$$

where $\eta_i = \eta_i(\boldsymbol{\beta}^*, \boldsymbol{\phi}^*)$ and $\bar{\eta}_n = n^{-1}\sum_{i=1}^n \eta_i$, satisfies

$$\frac{\hat{V}(\boldsymbol{\beta}^*, \boldsymbol{\phi}^*)}{V} \xrightarrow{p} 1,$$

where $V = n^{-1}\sigma^2$ and $\xrightarrow{p}$ denotes the convergence in probability. Therefore, by the Slutsky Theorem again, we have

$$\frac{\hat{\theta}_p - \theta}{\sqrt{\hat{V}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})}} \xrightarrow{L} N(0,1). \tag{3.19}$$

This asymptotic result can be used to construct confidence intervals for $\theta = E(Y)$. The reference distribution in (3.19) is either the OR model or the RP model.

The variance estimator (3.18) of the proposed DR estimator is computationally attractive because the linearized values, (3.16), are easy to compute.

4. Extension to Survey Sampling

We consider the problem of doubly robust inference in the survey sampling context. Consider a finite population U of size N. We are interested in estimating the mean of the finite population, $\theta_N = N^{-1}\sum_{i\in U} y_i$. To that end, a sample s, of size n is selected according to a given sampling design $p(s)$. In the complete data situation, a basic estimator is the expansion estimator given by (2.1) with $w_i = 1/(N\pi_i)$, where π_i denotes the first-order inclusion probability of unit i in the sample. In the presence of nonresponse to the y-variable, the imputed estimator $\hat{\theta}_p$ of θ_N is given by (2.3) with $w_i = 1/(N\pi_i)$. Note that $\hat{\theta}_p$ reduces to $\hat{\theta}_n$ in the complete data case.

In finite population sampling, the set of respondents can be viewed as the result of a three-stage process. First, the finite population is generated from an infinite population according to a given model. Then, a sample s of size n, is selected from the finite population according to a given sampling design $p(s)$. Finally, the set of respondents is generated from s according to the unknown nonresponse mechanism. Therefore, we identify three sources of randomness: the model m, which generates the vector of population values $\mathbf{Y}_U = (y_1, \ldots, y_N)'$;

the sampling design $p(s)$, which generates the vector of sample indicators $\mathbf{I}_U = (I_1, \ldots, I_N)'$ such that $I_i = 1$ if unit i is selected in the sample and $I_i = 0$, otherwise; the nonresponse mechanism, which generates the vector of response indicators $\boldsymbol{\delta}_U = (\delta_1, \ldots, \delta_N)'$. Here, the response indicator δ_i is defined for all the population units.

For the RP model approach, the vector $\mathbf{Y}_U$ is held fixed and, under the RP model approach, the properties of an estimator are evaluated under the joint distribution induced by the sampling design and the nonresponse mechanism. Given $\mathbf{Y}_U$, the population mean θ_N is a fixed quantity that we want to estimate.

For the OR approach, the properties of estimators are evaluated with respect to the joint distribution induced by the outcome regression model m and the sampling design. Here, the population mean θ_N is random so we face a prediction problem rather than an estimation problem. In both approaches, the vector $\mathbf{X}_U = (\mathbf{x}_1, \ldots, \mathbf{x}_N)'$ is held fixed.

We discuss the asymptotic properties of the DR estimator $\hat{\theta}_p$ of the form (2.3) using $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$, where $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$ is obtained by solving simultaneously (3.4) and (3.5). Under some regularity conditions, the asymptotic equivalence in (3.8) holds and the resulting imputed estimator is doubly robust.

Traditionally, the total variance of the DR estimator $\hat{\theta}_p$ has been expressed as the sum of the sampling variance and the nonresponse variance. This decomposition of the total variance results from viewing nonresponse as a second-phase of selection; e.g., Särndal (1992) and Deville and Särndal (1994), among others. We consider an alternative framework, which we call the reverse framework; e.g., Fay (1991), Rao and Shao (1992), Shao and Steel (1999) and Kim and Rao (2009). It consists of viewing the situation prevailing in the presence of nonresponse as follows: first, applying the nonresponse mechanism, the finite population U is randomly divided into a population of respondents U_r and a population of nonrespondents U_m; given (U_r, U_m), a sample s, containing both respondents and nonrespondents, is selected from U according to the given sampling design.

Under the RP model approach, the total variance of $\hat{\theta}_p$, $V(\hat{\theta}_p \mid \mathbf{X}_U, \mathbf{Y}_U)$, can be expressed as

$$V_T^{RP} = V_1^{RP} + V_2^{RP}, \tag{4.1}$$

where $V_1^{RP} = E\{V(\hat{\theta}_p \mid \mathbf{Y}_U, \mathbf{X}_U, \boldsymbol{\delta}_U)|\mathbf{Y}_U, \mathbf{X}_U\}$ and $V_2^{RP} = V\{E(\hat{\theta}_p \mid \mathbf{Y}_U, \mathbf{X}_U, \boldsymbol{\delta}_U)|\mathbf{Y}_U, \mathbf{X}_U\}$. Under the OR model, the total variance of $\hat{\theta}_p$ is

$$V_T = V_1^{OR} + V_2^{OR}, \tag{4.2}$$

where $V_1^{OR} = E\{V(\hat{\theta}_p - \theta_N \mid \mathbf{Y}_U, \mathbf{X}_U, \boldsymbol{\delta}_U) \mid \mathbf{X}_U, \boldsymbol{\delta}_U\}$ and $V_2^{OR} = V\{E(\hat{\theta}_p - \theta_N \mid \mathbf{Y}_U, \mathbf{X}_U, \boldsymbol{\delta}_U) \mid \mathbf{X}_U, \boldsymbol{\delta}_U\}$. An estimator of V_T^{RP} (respectively V_T^{OR}) is thus obtained by separately estimating V_1^{RP} and V_2^{RP} (respectively V_1^{OR} and V_2^{OR}). Under mild regularity conditions, the component V_1^{RP} (respectively V_1^{OR}) is of order $O\left(n^{-1}\right)$, whereas the components V_2^{RP} (respectively V_2^{OR}) is of order $O\left(N^{-1}\right)$. Therefore, the contribution of V_2^{RP} (respectively V_2^{OR}) to the total variance, V_2^{RP}/V_T^{RP} (respectively V_2^{OR}/V_T^{OR}) is of order $O\left(N^{-1}n\right)$ and is negligible when the sampling fraction n/N is negligible.

In order to estimate either V_1^{RP} or V_1^{OR}, it suffices to estimate $V(\hat{\theta}_p|\mathbf{Y}_U, \mathbf{X}_U, \boldsymbol{\delta}_U)$, the variance due to sampling conditional on $\mathbf{Y}_U, \mathbf{X}_U$ and $\boldsymbol{\delta}_U$. We can apply Theorem 1, which states that $\hat{\theta}_p$ is asymptotically equivalent to $\tilde{\theta}_p$ given by (3.9), so we can approximate $V(\hat{\theta}_p|\mathbf{Y}_U, \mathbf{X}_U, \boldsymbol{\delta}_U)$ by $V(\tilde{\theta}_p|\mathbf{Y}_U, \mathbf{X}_U, \boldsymbol{\delta}_U)$. For example, for a fixed size or random size without replacement sampling design, we have

$$V(\tilde{\theta}_p|\mathbf{Y}_U, \mathbf{X}_U, \boldsymbol{\delta}_U) = \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U}(\pi_{ij} - \pi_i\pi_j)\,\frac{\eta_i}{\pi_i}\frac{\eta_j}{\pi_j}, \tag{4.3}$$

where η_i is given by (3.16) and π_{ij} denotes the second order inclusion probability for units i and j. An estimator of V_1^{RP} (respectively V_1^{OR}), denoted by $\hat{V}_1$, is then

$$\hat{V}_1 = \frac{1}{N^2}\sum_{i\in s}\sum_{j\in s}\frac{(\pi_{ij} - \pi_i\pi_j)}{\pi_{ij}}\frac{\hat{\eta}_i}{\pi_i}\frac{\hat{\eta}_j}{\pi_j},$$

where $\hat{\eta}_i$ is obtained from η_i by replacing $(\boldsymbol{\beta}_0, \boldsymbol{\phi}_0)$ with $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$. Note that $\hat{V}_1$ is obtained by applying a complete data variance estimation method to $\hat{\eta}_i$ in the sample. Under mild regularity conditions (e.g., Deville (1999)), the estimator $\hat{V}_1$ is consistent for either V_1^{RP} or V_1^{OR} regardless of the validity of the assumed RP or OR model. Consistency of $\hat{V}_1$ follows from standard regularity conditions used in the complete data case. If the sampling fraction n/N is negligible, a consistent estimator of the total variance of $\hat{\theta}_p$ (under either the RP or the OR model) is given by $\hat{V}_1$.

When the sampling fraction is not negligible, one must take the term V_2^{RP} into account (in the case of the RP model) or V_2^{OR} (in the case of the OR model). Once again, we use the asymptotic equivalence between $\hat{\theta}_p$ and $\tilde{\theta}_p$ established in Theorem 1. We have

$$E(\tilde{\theta}_p - \theta_N|\mathbf{Y}_U, \mathbf{X}_U, \boldsymbol{\delta}_U) = \frac{1}{N}\sum_{i\in U}(\eta_i^* - y_i),$$

where $\eta_i^* = \eta_i(\boldsymbol{\beta}^*, \boldsymbol{\phi}^*)$ is defined as (3.16). Under the RP model,

$$\begin{aligned}V_2^{RP} &= V\left\{E(\tilde{\theta}_p - \theta_N|\mathbf{Y}_U, \mathbf{X}_U, \boldsymbol{\delta}_U)|\mathbf{Y}_U, \mathbf{X}_U\right\}\\ &= \frac{1}{N^2}\sum_{i\in U}\frac{p_i(1-p_i)}{p_i^2}\{y_i - m(x_i, \boldsymbol{\beta}^*)\}^2.\end{aligned}$$

Thus, an estimator of V_2^{RP}, denoted by $\hat{V}_2$, is

$$\hat{V}_2 = \frac{1}{N^2}\sum_{i\in s}\pi_i^{-1}\delta_i\frac{(1-p_i(\hat{\boldsymbol{\phi}}))}{p_i(\hat{\boldsymbol{\phi}})^2}\hat{e}_i^2, \tag{4.4}$$

where $\hat{e}_i = y_i - m(x_i, \hat{\boldsymbol{\beta}})$. Because $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$ is a consistent estimator of $(\boldsymbol{\beta}^*, \boldsymbol{\phi}_0)$ under the RP model, $\hat{V}_2$ in (4.4) is asymptotically unbiased and consistent for V_2^{RP} under the RP model. Therefore, a consistent estimator of the total variance under the RP model is given by

$$\hat{V}_T = \hat{V}_1 + \hat{V}_2. \tag{4.5}$$

To see if $\hat{V}_T$ in (4.5) is doubly robust, one needs to check if $\hat{V}_2$ in (4.4) is consistent for V_2^{OR} under the OR model. We first note that

$$\begin{aligned}
V_2^{OR} &= V\left\{E(\tilde{\theta}_p - \theta_N|\mathbf{Y}_U, \mathbf{X}_U, \boldsymbol{\delta}_U) \mid \mathbf{X}_U, \boldsymbol{\delta}_U\right\} \\
&= \frac{1}{N^2}\sum_{i\in U}\left(\frac{\delta_i}{p_i(\boldsymbol{\phi}^*)} - 1\right)^2 V(y_i \mid \mathbf{x}_i) \\
&= \frac{1}{N^2}\sum_{i\in U}\left\{\frac{\delta_i}{p_i(\boldsymbol{\phi}^*)^2} - \frac{2\delta_i}{p_i(\boldsymbol{\phi}^*)} + 1\right\}V(y_i \mid \mathbf{x}_i).
\end{aligned} \tag{4.6}$$

Thus, the asymptotic bias of $\hat{V}_2$ in (4.4), as an estimator of V_2^{OR} under the OR model, is

$$E\left\{\hat{V}_2\right\} - V_2^{OR} \doteq \frac{1}{N^2}\sum_{i\in U}E\left\{\frac{\delta_i}{p_i(\boldsymbol{\phi}^*)} - 1\right\}V(y_i \mid \mathbf{x}_i). \tag{4.7}$$

Thus, under the OR model, if we further assume that $V(y_i \mid \mathbf{x}_i) = \psi(\mathbf{x}_i; \alpha_0)$ for some α_0 and a consistent estimator $\hat{\alpha}$ is available, then the right side of (4.7) can be estimated by

$$\hat{B}\left(\hat{V}_2\right) = \frac{1}{N^2}\sum_{i\in s}\pi_i^{-1}\left\{\frac{\delta_i}{p_i(\hat{\boldsymbol{\phi}})} - 1\right\}\psi(\mathbf{x}_i; \hat{\alpha}). \tag{4.8}$$

The expected value of the estimated bias term in (4.8) is asymptotically equal to zero under the RP model because $p_i(\hat{\boldsymbol{\phi}})$ converges to the true response probability. Also, we expect the term $\hat{B}(\hat{V}_2)$ to be large if the RP model is misspecified and the OR model does not fit the data well, in which case the quantity $\psi(\mathbf{x}_i; \hat{\alpha})$ is likely to be large. Thus, the bias-adjusted estimator of the total variance

$$\hat{V}_T = \hat{V}_1 + \hat{V}_2 - \hat{B}(\hat{V}_2) \tag{4.9}$$

is doubly robust.

5. Simulation Study

We performed two simulation studies. The first, presented in Section 5.1, compares the performance of several point and variance estimators in the infinite population set-up. In Section 5.2, the case of finite population sampling is considered.

5.1. Infinite population set-up

The simulation study can be described as a $2 \times 2 \times 5$ factorial design with $R = 5{,}000$ replications within each cell. The factors are two types of sampling distributions, two types of the nonresponse mechanisms, and five types of point estimators. For the sampling distributions, the first was generated from a linear regression model, and the second was generated according to a non-linear model. For the linear model, we used

$$y_i = 1 + x_{1i} + \epsilon_i, \tag{5.1}$$

where $x_{1i} \sim N(1,1)$, $\epsilon_i \sim N(0,1)$, and x_{1i} and ϵ_i are independent. For the non-linear model, we used the same x_{1i} and ϵ_i, but y_i was generated independently according to

$$y_i = 0.5(x_{1i} - 1.5)^2 + \epsilon_i. \tag{5.2}$$

Two random samples of size $n = 500$ were separately generated from the two models. From each sample, we generated two types of the respondents from $Bernoulli(p_{1i})$ (Type A) and $Bernoulli(p_{2i})$ (Type B), respectively, with $\text{logit}(p_{1i}) = x_{2i}$ and $\text{logit}\,(p_{2i}) = -0.5 + 0.5(x_{2i} - 2)^2$, where $x_{2i} \sim \exp(1)$ and x_{2i} is independent of (x_{1i}, ϵ_i). The overall response rates were about 60% in both cases.

In each sample, we computed five estimators for $\theta = E(Y)$: the complete sample estimator ($\hat{\theta}_n = n^{-1}\sum_{i=1}^{n} y_i$, Complete); The proposed doubly robust estimator, (New); the doubly robust estimator of Haziza and Rao (2006), (HR); the doubly robust estimator of Cao, Tsiatis, and Davidian (2009), (CTD); the doubly robust estimator of Tan (2006), (Tan).

We considered three scenarios at the estimation stage:

1. Scenario 1: Both models are correct, the sample was generated from (5.1) and the respondents were generated from the Type A model. The "working" OR model is $E(y_i \mid x_{1i}) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_{1i}$ and the "working" RP model is $\delta_i \sim Bernoulli(p_i)$ with $\text{logit}(p_i) = \phi_0 + \phi_1 x_{2i}$.
2. Scenario 2: Only the OR model is correct, we used the working models in Scenario 1 but the sample was generated from (5.1) and the respondents were generated from the Type B model.

3. Scenario 3: Only the RP model is correct, we used the working models in Scenario 1 but the sample was generated from (5.2) and the respondents were generated from the Type A model.

For the estimators HR, CTD and Tan, $(\hat{\phi}_0, \hat{\phi}_1)$ was computed by maximum likelihood, whereas it was computed by solving

$$\sum_{i=1}^{n} w_i \frac{\delta_i}{p_i(\phi)} (1, x_{2i}) = \sum_{i=1}^{n} w_i (1, x_{2i}) \tag{5.3}$$

for the New estimator, where $\phi = (\phi_0, \phi_1)$ and $w_i = 1/n$. Once the $\hat{p}_i$'s were computed, both HR and the New methods used $(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_1)$ given by

$$(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_1)' = \left\{ \sum_{i=1}^{n} w_i \delta_i \left(\hat{p}_i^{-1} - 1\right) \mathbf{x}_i \mathbf{x}_i' \right\}^{-1} \sum_{i=1}^{n} w_i \delta_i \left(\hat{p}_i^{-1} - 1\right) \mathbf{x}_i y_i, \tag{5.4}$$

where $\mathbf{x}_i = (1, x_{1i})'$. For the CTD estimator, we used

$$(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_1, \hat{c}_0, \hat{c}_1)' = \left\{ \sum_{i=1}^{n} w_i \delta_i \hat{p}_i^{-1} \left(\hat{p}_i^{-1} - 1\right) \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i' \right\}^{-1} \sum_{i=1}^{n} w_i \delta_i \hat{p}_i^{-1} \left(\hat{p}_i^{-1} - 1\right) \tilde{\mathbf{x}}_i y_i, \tag{5.5}$$

where $\tilde{\mathbf{x}}_i = (1, x_{1i}, \hat{p}_i, \hat{p}_i x_{2i})'$. The doubly robust estimator of Tan (2006) is computed as

$$\hat{\theta}_{tan} = \sum_{i=1}^{n} w_i \frac{\delta_i y_i}{\hat{p}_i} - \sum_{i=1}^{n} w_i \Big(\frac{\delta_i}{\hat{p}_i} - 1 \Big) \Big(\hat{k}_0 + \hat{k}_1 \hat{m}_i \Big),$$

where $\hat{m}_i = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 x_{1i}$ and

$$(\hat{k}_0, \hat{k}_1, \hat{d}_0, \hat{d}_1)' = \left\{ \sum_{i=1}^{n} w_i \delta_i \hat{p}_i^{-1} \left(\hat{p}_i^{-1} - 1\right) \tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i' \right\}^{-1} \sum_{i=1}^{n} w_i \delta_i \hat{p}_i^{-1} \left(\hat{p}_i^{-1} - 1\right) \tilde{\mathbf{z}}_i y_i, \tag{5.6}$$

where $\tilde{\mathbf{z}}_i = (1, \hat{m}_i, \hat{p}_i, \hat{p}_i x_{2i})'$.

Table 1 presents the Monte Carlo averages and variances of the five estimators under the different scenarios. New, HR, CTD, and Tan were all approximately unbiased in all scenarios, illustrating that their double robustness. Turning to relative efficiency, the New estimator showed the best performances in all cases. In Scenario 1, the CTD estimator had the largest variance. In Scenario 2, the New estimator showed the best performance - the calibration condition (5.3) can be justified as the optimality condition when the OR model is true. Tan's estimator showed slightly higher variance under Scenario 2, whereas the CTD estimator had slightly higher variance under Scenario 3.

Table 1. Monte Carlo average and variance of the point estimators in simulation one.

Scenario	Method	Mean	Variance	Standardized Variance
	Complete	2.00	0.003925	100
1	New	2.00	0.005524	141
(Both models true)	CTD	2.00	0.005907	150
	HR	2.00	0.005524	141
	Tan	2.00	0.005530	141
	Complete	2.00	0.003925	100
2	New	2.00	0.005278	134
(OR model true)	CTD	2.00	0.005287	135
	HR	2.00	0.005360	137
	Tan	2.00	0.005623	143
	Complete	0.62	0.003466	100
3	New	0.62	0.005936	171
(RP model true)	CTD	0.62	0.006540	189
	HR	0.62	0.005939	171
	Tan	0.62	0.005942	171

We only consider variance estimation for the CTD method and the New method. The variance estimator proposed by Cao, Tsiatis, and Davidian (2009) was computed using (3.18) with

$$\eta_i = m(\mathbf{x}_i;\hat{\boldsymbol{\beta}}) + \frac{\delta_i}{\hat{p}_i}\left\{y_i - m(\mathbf{x}_i;\hat{\boldsymbol{\beta}})\right\} - \hat{\mathbf{c}}\,(\delta_i - \hat{p}_i)\,(1, x_{2i})', \tag{5.7}$$

where $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_1)$ and $\hat{\mathbf{c}} = (\hat{c}_0, \hat{c}_1)$ were computed from (5.5). The variance estimator for the New estimator was computed using (3.18) with

$$\eta_i = m(\mathbf{x}_i;\hat{\boldsymbol{\beta}}) + \frac{\delta_i}{\hat{p}_i}\left\{y_i - m(\mathbf{x}_i;\hat{\boldsymbol{\beta}})\right\} \tag{5.8}$$

and $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_1)$ given by (5.4). In (5.8), we obtained $\hat{p}_i$ using maximum likelihood. Variance estimation in the context of Tan's estimator was not computed here as Tan (2006) did not discuss variance estimation.

Table 2 gives the Monte Carlo bias of the variance estimators and the coverage of the interval estimators of the CTD and the New estimators. We used $(\hat{\theta} - 1.96\sqrt{\hat{V}}, \hat{\theta} + 1.96\sqrt{\hat{V}})$ for interval estimation. The proposed variance estimator for the New estimator showed small relative biases (less than 5% in absolute values) in all scenarios, suggesting that the variance estimator for the New estimator is doubly robust. The variance estimator for CTD showed somewhat modest bias (8.27%) under Scenario 3.

Table 2. Monte Carlo percent relative bias of the two variance estimators and coverage of the two interval estimators in simulation one.

Scenario	Method	Relative Bias (%)	Coverage (%)
1	New	2.27	95.1
	CTD	5.75	94.9
2	New	4.69	95.5
	CTD	3.33	95.4
3	New	-0.04	94.7
	CTD	8.27	94.7

Table 3. Characteristics of the population.

Stratum	1	2	3	4
N_h	2000	1500	1000	500
β_{0h}	10	15	20	25
β_{1h}	1	1.5	2	2.5
β_{2h}	1	1.5	2	2.5
R_h^2	0.64	0.52	0.57	0.61

5.2. Survey sampling set-up

We carried out a simulation study in the survey sampling set-up. We generated a population of size $N = 5,000$ consisting of four strata $U_1, \ldots, U_4$ of size $N_1, \ldots, N_4$, respectively. We generated 5 variables: a variable of interest y and three auxiliary variables x_1-x_3. Each of the x-variable was independently generated from a Gamma distribution with parameters 2 and 25. Then, given x_1 and x_2, the y-values were generated according to

$$y_i = \beta_{0h} + \beta_{0h} + \beta_{1h}x_{1i} + \beta_{2h}x_{2i} + \epsilon_i, \qquad \text{for } i \in U_h,$$

where the ϵ_i's were generated from a normal distribution with mean 0 and variance σ_h^2, $i \in U_h$, whose value was set to lead to a given coefficient of determination (R_h^2). The characteristics of the population are shown in Table 3.

The objective consisted in estimating the finite population mean $\theta_N = n^{-1}\sum_{i \in U} y_i$. From the population we generated $R = 5,000$ samples, in each stratum a simple random sample s_h of size n_h was selected from U_h, $h = 1, 2, 3, 4$. Equal allocation was used with $n_h = 125$ and $n_h = 250$, which correspond to an overall sampling fraction of 10% and 20%, respectively. This particular design leads to unequal probability of selection for units in different strata.

In each selected sample, nonresponse to the study variable y was generated according to

$$\text{logit } (p_i) = -2 + 0.03x_{1i} + 0.03x_{2i}.$$

In each stratum, the response rate was approximately equal to 70%.

We computed five estimators of the mean: the complete sample estimator (C) given by (2.1) with $w_i = 1/(N\pi_i)$; the propensity score adjusted (PSA) estimator given by (2.1) with $w_i = 1/(N\pi_i\hat{p}_i)$; the estimator of Haziza and Rao (2006) (HR); the estimator of Cao, Tsiatis, and Davidian (2009) (CTD); the proposed estimator (New). We considered three scenarios:

(i) Scenarios 1: The RP and the OR models were correctly specified.

(ii) Scenario 2: Only the OR model was correctly specified. For the RP working model, we used $\text{logit}(p_i) = \phi_0 + \phi_1 x_{1i} + \phi_2 x_{3i}$.

(iii) Scenario 3: Only the RP model was correctly specified. For the OR working model, we used $E(y_i \mid \mathbf{x}_i) = \beta_0 + \beta_1 x_{1i} + \beta_3 x_{3i}$.

As a measure of the bias of an estimator $\hat{\theta}$, we used the Monte Carlo Percent Relative Bias (RB),

$$RB(\hat{\theta}) = 100 \times \frac{E_{MC}(\hat{\theta}) - \theta_N}{\theta_N},$$

where $E_{MC}(\hat{\theta}) = R^{-1}\sum_{r=1}^{R}\hat{\theta}^{(r)}$ and $\hat{\theta}^{(r)}$ denotes the estimator $\hat{\theta}$ for the r-th sample. To compare the efficiency of the estimation procedures, we computed the percent relative efficiency, using the complete sample estimator as the reference.

In each sample, we computed the estimator of the total variance (corresponding to the New estimator) given by (4.9). In order to compute (4.8), we used

$$\psi(\mathbf{x}_i; \hat{\alpha}) = \frac{\sum_{i\in s} w_i \delta_i \hat{e}_i^2}{\sum_{i\in s} w_i \delta_i},$$

where $\hat{e}_i$ denotes the residual attached to unit i obtained after fitting the working outcome regression model. As a measure of the bias of $\hat{V}_T$, we used the Monte Carlo percent relative bias of the variance estimator. The relative bias of $\hat{V}_T$ is shown in Table 4 (in parentheses).

Table 4 presents the Monte Carlo percent relative bias and percent relative efficiency (with respect to the complete data estimator) of five estimators under the three scenarios. The HR, CTD, and New estimators all showed negligible bias in all scenarios, which is an indication that they are all doubly robust. The PSA estimator showed a modest bias when the RP model was misspecified. In terms of efficiency, the estimators HR, CTD, and New showed similar performances in all the scenarios, although the CTD was slightly less efficient than the other two. When the RP model was misspecified (Scenario 2), the PSA estimator showed a low efficiency, as expected, and the other two estimators showed almost identical performances. In Table 4, the proposed variance estimator shows good performances in all scenarios (with a relative absolute bias less than 5%).

Table 4. Monte Carlo percent relative bias and percent relative efficiency of five estimators and Monte Carlo percent relative bias of the proposed variance estimator.

		$f = 0.1$		$f = 0.2$	
Scenario	Method	RB	RE	RB	RE
1	Complete	0.00	100	0.00	100
	PSA	0.04	271	0.01	278
	HR	0.04	270	0.01	278
	CTD	0.08	274	0.03	281
	New	0.04	270	0.01	278
		(-1.2)		(2.5)	
2	Complete	0.00	100	0.00	100
	PSA	0.76	394	0.75	550
	HR	0.03	267	0.01	296
	CTD	0.05	270	0.02	298
	New	0.03	267	0.01	296
		(-1.4)		(-3.1)	
3	Complete	0.00	100	0.01	100
	PSA	0.05	265	0.04	288
	HR	0.06	264	0.05	289
	CTD	0.09	268	0.06	291
	New	0.06	264	0.05	288
		(3.7)		(1.7)	

Percent relative biases of the variance estimators are in parenthesis.

6. Concluding remarks

In this paper, we proposed a new doubly robust estimator that showed good finite sample performances in simulation studies. The resulting variance estimator is also doubly robust and can be readily implemented using complete data software.

The proposed estimator is based on single deterministic imputation and it is well known that the single imputation of the form (2.3) can lead to biased estimates for the population proportions. In this case, fractional imputation, considered by Fay (1996), Kim and Fuller (2004) and Fuller and Kim (2005), can be used to obtain valid estimates for several parameters. Doubly robust fractional imputation will be discussed elsewhere.

In the simulation studies, the new method showed better efficiency than the other doubly robust estimators in most cases, but there is no guarantee that it is uniformly optimal. Further investigation in this direction is also a topic of future research.

Acknowledgement

We thank professor J. N. K. Rao, three anonymous referees and an Associate Editor for their constructive comments. The research of the first author was partially supported by a Cooperative Agreement between the US Department of Agriculture Natural Resources Conservation Service and Iowa State University. The research of the second author was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

References

Bang, H. and Robins, J. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 962-973.

Chang, T. and Kott, P. S. (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika* **95**, 555-571.

Cao, W., Tsiatis, A. A. and Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*, **96**, 723-734.

Cochran, W. G. (1977). *Sampling Techniques.* Wiley, New York.

Deville, J.-C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Surv. Methodol.* **25**, 193-203.

Deville, J. C. and Särndal, C. E. (1994). Variance estimation for the regression imputed Horvitz-Thompson estimator. *J. Official Statist.* **10**, 381-394.

Fay, R. E. (1991). A Design-Based Perspective on Missing Data Variance. *Proceedings of the* 1991 *Annual Research Conference, US Bureau of the Census*, 429-440.

Fay, R. E. (1996). Alternative paradigms for the analysis of imputed survey data. *J. Amer. Statist. Assoc.* **91**, 490-498.

Firth, D. and Bennett, K. E. (1998). Robust models in probability sampling. *J. Roy. Statist. Soc. Ser. B* **60**, 3-21.

Folsom, R. E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction. *Proceedings of the Social Statistics Section*, 197-202. American Statistical Association.

Folsom, R. E. and Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. *Proceedings of the Survey Research Methods Section,* 596-603. American Statistical Association.

Fuller, W. A. (2009). *Sampling Statistics.* Wiley, Hoboken, New Jersey.

Fuller, W. A. and Kim, J. K. (2005). Hot deck imputation for the response model, *Surv. Methodol.* **31**, 139-149.

Haziza, D. and Rao, J. N. K. (2006). A nonresponse model approach to inference under imputation for missing survey data. *Surv. Methodol.* **32**, 53-64.

Iannacchione, V. G., Milne, J.G., and Folsom, R.E. (1991). Response probability weight adjustments using logistic regression. *Proceeding in Survey Research Method Section*, 637-642. American Statistical Association.

Kang, J. D. Y. and Schafer, J. L. (2008). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statist. Sci.* **22**, 523-539.

Kim, J. K. and Fuller, W. A. (2004). Fractional hot deck imputation. *Biometrika* **91**, 559-578.

Kim, J. K. and Kim, J. J. (2007). Nonresponse weighting adjustment using estimated response probability. *Canad. J. Statist.* **35**, 501-514.

Kim, J. K. and Park, H. A. (2006). Imputation using response probability. *Canad. J. Statist.* **34**, 171-182.

Kim, J.K. and Rao, J.N.K. (2009). Unified approach to linearization variance estimation from survey data after imputation for item nonresponse. *Biometrika* **96**, 917-932.

Kim, J. K. and Riddles, M. (2012). Some theory for propensity scoring adjustment estimator. *Surv. Methodol.* **38**, 157-165.

Kott, P. S. (1994). A note on handling nonresponse in sample surveys. *J. Amer. Statist. Assoc.* **89**, 693-696.

Kott, P.S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Surv. Methodol.* **32**, 133-142.

Kott, P.S. and Chang, T. (2010). Using calibration weighting to adjust for nonignorable unit nonresponse. *J. Amer. Statist. Assoc.* **105**, 1265-1275.

Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *Ann. Statist.* **10**, 462-474.

Rao, J. N. K. and Shao, J. (1992). Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* **79**, 811-822.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficient when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89**, 846-866.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581-592.

Rubin, D. B. and van der Laan, M. J. (2008). Empirical efficiency maximization: Improved locally efficient covariate adjustment in randomized experiments and survival analysis. *International Journal of Biostatistics* **4**.

Särndal, C. E. (1992). Method for estimating the precision of survey estimates when imputation has been used. *Survey Methodology* **18**, 241-252.

Särndal, C. E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling.* Springer-Verlag.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion and rejoinder). *J. Amer. Statist. Assoc.* **94**, 1096-1146.

Shao, J. and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *J. Amer. Statist. Assoc.* **94**, 254-265.

Tan, Z. (2006). A distributional approach for causal inference using propensity scores. *J. Amer. Statist. Assoc.* **101**, 1619-1637.

Department of Statistics, Iowa State University, Ames, Iowa, 50011, U.S.A.

E-mail: jkim@iastate.edu

Département de mathématiques et de statistique, Université de Montréal, Montreal, H3C 3J7, Canada.

E-mail: david.haziza@umontreal.ca

(Received January 2012; accepted December 2012)