Efficient nonresponse weighting adjustment using estimated response probability

Jae Kwang Kim
Department of Applied Statistics, Yonsei University, Seoul, 120-749, KOREA

Key Words: Regression estimator, Propensity score, Survey sampling, Unit Nonresponse, Variance estimation.

1. Introduction

Weighting adjustment is a popular method of handling unit nonresponse in sample surveys. Brick and Kalton (1996) and Groves et al. (2002) provide comprehensive overview of the nonresponse weighting adjustment (NWA) methods in survey sampling. For the respondents to properly represent the original population, the increase of the original weight of the respondent is carefully determined based on the auxiliary variables observed throughout the sample.

Using a theory for two-phase sampling, where the set of the respondents is treated as a second phase sample from the original sample, the selection bias for the respondents can be safely removed by inversely multiplying the response probability. Since the true response probabilities are usually unknown, as in Rosenbaum (1987), Ekholm and Laaksonen (1991), and Iannacchione (2003), estimated probabilities can be used to correct for the nonresponse bias. Since the response probabilities are directly used, the method is called the direct NWA method. The direct NWA estimator is mainly motivated to reduce the nonresponse bias.

When the auxiliary variables that are related with the study variable of interest are also available, we can improve the efficiency of the direct NWA estimator by incorporating the auxiliary variables into the estimation procedure. Regression weighting method is one of the commonly used technique for incorporating the auxiliary variable in this situation. The regression NWA method modifies the weights of the direct NWA estimator in such a way that the weighted sums of the auxiliary variables for the respondents using the modified weights are equal to the weighted sums of the auxiliary variables for the original samples using the original weights, respectively. Bethlehem (1988), Fuller and An (1998), and Lundström and Särndal (1999) discussed the regression NWA method.

Another commonly used method, called weighting cell adjustment, should be also mentioned. Weighting cell adjustment increases the weights of the respondents by the same amount in each cell so that the sum of the adjusted weights of the respondents reproduces the sum of the original weights of the whole sample within each cell. A weighting cell consists of sample units which are believed to have approximately equal response probabilities or to have similar item values. Under the equal response probability assumption, the method using weighting cell is a special case of the direct NWA estimator. Also, under the homogeneity of the study item assumption, the cell weighting estimator is a special case of the regression NWA estimator. Little (1986) provided an overview of the weighting cell method, Eltinge and Yansaneh (1997) discussed diagnostics for the formation of cells, and Rizzo et al (1996) compared several weighting adjustment methods using a data from the Survey of Income and Program Participation.

The NWA estimators discussed above reduces the nonresponse biases because the estimated response probability is incorporated into the estimator. Although commonly used in practice, asymptotic properties of the NWA estimators using the estimated response probability are not fully discussed in the literature. In particular, Rosenbaum (1987) and Robins et al (1994) noted that the estimator using the estimated response probability can be more efficient than the estimator using the true response probability, but they did not fully explain when and why this phenomenon holds. In this paper, under model for the response fairly general parametric model for the response probability, we show that the estimator using the estimated response probability is always no less efficient than the estimator using true response probability. The implication is that, even when we know the true response probability, it is always better to use its estimate for the NWA weighting. Also, we propose an estimator for the variance of the NWA estimator using the estimated response probability.

The paper is organized as follows. In Section 2, we introduce the notation and discuss the properties of the direct NWA estimators. In Section 3, the NWA estimator using the regression weighting method is discussed. In Section 4, variance estimation is discussed and results from a limited simulation study are presented in Section 5.

2. Direct NWA estimation

Let the finite population be of size N, indexed from 1 to N, and N is assumed to be known. Let the parameter of interest be the population mean $\bar{y}_N = N^{-1}\sum_{i=1}^N y_i$, where y_i is the study variables for unit i. Let $\bar{y}_n$ be an estimator of $\bar{y}_N$ based on the sample of size n and of the form $\bar{y}_n = \sum_{i\in A} w_i y_i$, where w_i is the weight of unit i satisfying $\sum_{i\in A} w_i = 1$ and A is the set of indices in the sample. We assume that

$$E\left(\bar{y}_n \mid \mathcal{F}_N\right) = \bar{y}_N, \tag{1}$$

where $\mathcal{F}_N = \{y_1, y_2, \cdots, y_N\}$ and the expectation in (1) is taken with respect to the sampling mechanism.

Under nonresponse, we define the response indicator function of y_i

$$R_i = \begin{cases} 1 & y_i \text{ responds} \\ 0 & y_i \text{ does not respond} \end{cases}, \quad i \in A,$$

and let $\mathcal{R} = \{(i, R_i) : i \in A\}$ be the set of response indicators for all units in the sample. Let $\pi_i = \Pr\left(R_i = 1 \mid i \in A\right)$ be the response probability of sampled unit i. If we know the response probability π_i, then the population mean is unbiasedly estimated by multiplying the inverse of π_i to the original weight. Let

$$\bar{y}_d = \frac{\sum_{i\in A} w_i \pi_i^{-1} R_i y_i}{\sum_{i\in A} w_i \pi_i^{-1} R_i} \tag{2}$$

be such an estimator constructed by directly using the true response probability. The denominator term in (2) is included so that the final weights add up to one. By the definition of π_i, the conditional expectation of $\bar{y}_d$, conditional on A, approximately equals to $\bar{y}_n$ and, hence, $\bar{y}_d$ is approximately unbiased for $\bar{y}_N$.

When the true response probability π_i is not available, we use its estimate $\hat{\pi}_i$ using a suitable model for the response probability. Let

$$\bar{y}_e = \frac{\sum_{i\in A} w_i \hat{\pi}_i^{-1} R_i y_i}{\sum_{i\in A} w_i \hat{\pi}_i^{-1} R_i} \tag{3}$$

be an estimator of the population total using the estimated response probability $\hat{\pi}_i$. The estimators in (3) is called the direct NWA estimator because we directly use the estimated response probability and no other auxiliary variables are incorporated. The estimator $\bar{y}_d$ in (2) is not applicable in practice because the true response probabilities are not usually known.

We assume that the response probability for unit i can be written

$$\pi_i = \pi\left(\mathbf{z}_i; \boldsymbol{\alpha}^0\right), \tag{4}$$

for some known function $\pi\left(\mathbf{z}_i; \cdot\right)$ of parameter $\boldsymbol{\alpha}$ evaluated at $\boldsymbol{\alpha} = \boldsymbol{\alpha}^0$, where $\mathbf{z}_i$ is a vector of variables that can be observed for both respondents and nonrespondents. We assume that the true value $\boldsymbol{\alpha}^0$ is estimated by $\hat{\boldsymbol{\alpha}}$, which is the unique solution to the maximum likelihood equation:

$$\frac{\partial}{\partial\boldsymbol{\alpha}} \sum_{i\in A}\left[R_i \ln\left(\pi_i\right) + \left(1 - R_i\right)\ln\left(1 - \pi_i\right)\right] = 0. \tag{5}$$

If we define $H\left(\mathbf{z}_i; \boldsymbol{\alpha}\right) = \partial\left\{\text{logit}\left(\pi_i\right)\right\}/\partial\boldsymbol{\alpha}$, the maximum likelihood equation (5) can be written

$$\sum_{i\in A}\left(R_i - \pi_i\right) H\left(\mathbf{z}_i; \boldsymbol{\alpha}\right) = 0$$

and, under the regularity conditions, the resulting estimator $\hat{\boldsymbol{\alpha}}$ satisfies, conditional on A,

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0 = I\left(\boldsymbol{\alpha}^0\right)^{-1} \sum_{i\in A}\left(R_i - \pi_i\right) H\left(\mathbf{z}_i; \boldsymbol{\alpha}^0\right) + o_p\left(n^{-1/2}\right), \tag{6}$$

where $I\left(\boldsymbol{\alpha}^0\right)$ is the expected information matrix defined by

$$I\left(\boldsymbol{\alpha}^0\right) = E\left\{\sum_{i\in A} \pi_i\left(1 - \pi_i\right) H\left(\mathbf{z}_i; \boldsymbol{\alpha}^0\right) H\left(\mathbf{z}_i; \boldsymbol{\alpha}^0\right)' \mid \mathcal{F}_N\right\}.$$

To study the asymptotic properties of $\bar{y}_e$, we apply a Taylor expansion to get

$$\hat{\pi}_i^{-1} = \pi_i^{-1} + \left(\partial\pi_i^{-1}/\partial\boldsymbol{\alpha}\right)'\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0\right) + O_p\left(n^{-1}\right). \tag{7}$$

Since $\partial\pi_i^{-1}/\partial\boldsymbol{\alpha} = -\pi_i^{-2}\partial\pi_i/\partial\boldsymbol{\alpha} = -\left(\pi_i^{-1} - 1\right) H\left(\mathbf{z}_i; \boldsymbol{\alpha}^0\right)$, we have

$$\bar{y}_e = \bar{y}_d - \Delta_{yz}'\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0\right) + O_p\left(n^{-1}\right), \tag{8}$$

where $\Delta_{yz} = E\left\{\sum_{i\in A} w_i\left(1 - \pi_i\right)\left(y_i - \bar{y}_n\right) H\left(\mathbf{z}_i; \boldsymbol{\alpha}^0\right) \mid \mathcal{F}_N\right\}$. Since the Δ_{yz} corresponds to the population covariance between y_i and $\mathbf{z}_i$ among nonrespondents, Δ_{yz} is bounded. By (5), the second term in the right side of equality (8) is of order $n^{-1/2}$ with zero expectation. Hence, $\sqrt{n}$ consistency of $\bar{y}_e$ directly follows.

To investigate the variance of $\bar{y}_e$, note that

$$Var\begin{pmatrix} \bar{y}_d \\ \hat{\boldsymbol{\alpha}} \end{pmatrix} \mid \mathcal{F}_N \Bigg) = \begin{pmatrix} V_{11} & V_{12} \\ V_{12}' & V_{22} \end{pmatrix}$$

where, ignoring $o\left(n^{-1}\right)$ terms,

$$\begin{aligned} V_{11} &= Var\left(\bar{y}_n \mid \mathcal{F}_N\right) \\ &\quad + E\left\{\sum_{i\in A} w_i^2\left(\pi_i^{-1}-1\right)\left(y_i-\bar{y}_N\right)^2 \mid \mathcal{F}_N\right\} \\ V_{12} &= Cov\left(\bar{y}_d, \hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0 \mid \mathcal{F}_N\right) \\ &= \Delta_{yz}' I\left(\boldsymbol{\alpha}^0\right)^{-1} \end{aligned}$$

and

$$V_{22} = Var\left(\hat{\boldsymbol{\alpha}}\right) = I\left(\boldsymbol{\alpha}^0\right)^{-1}.$$

Thus, we have $\Delta_{yz}' = V_{12}V_{22}^{-1}$ and, ignoring the $o\left(n^{-1}\right)$ terms,

$$\begin{aligned} Var\left(\bar{y}_e \mid \mathcal{F}_N\right) &= Var\left(\bar{y}_d \mid \mathcal{F}_N\right) - \Delta_{yz}'V_{22}\Delta_{yz} \\ &\leq Var\left(\bar{y}_d \mid \mathcal{F}_N\right). \end{aligned} \tag{9}$$

The result in (9) shows that the efficiency of $\bar{y}_d$ can be improved if we use the estimated response probability. The variances of the two estimators, $\bar{y}_e$ and $\bar{y}_d$, are approximately the same if the Δ_{yz} term is approximately zero, as is the case when the auxiliary variable $\mathbf{z}_i$ is independent of the study variable y_i in the population.

3. Regression NWA estimation

So far, we have assumed the existence of the auxiliary variable $\mathbf{z}_i$ that is related with the response probability π_i. Now suppose that, in addition to $\mathbf{z}_i$, there is another auxiliary variable $\mathbf{x}_i$ that is related with the study variable y_i. If the auxiliary variable $\mathbf{x}_i$ are observed throughout the sample, a version of regression weighting method can be used. For a general description of the regression weighting method in the standard survey sampling setup, see Fuller (2002). The regression weighting estimator under nonresponse can be constructed as

$$\bar{y}_{re} = \bar{\mathbf{x}}_n'\hat{\boldsymbol{\beta}}_e \tag{10}$$

where

$$\begin{aligned} \bar{\mathbf{x}}_n &= \sum_{i\in A} w_i\mathbf{x}_i \\ \hat{\boldsymbol{\beta}}_e &= \left(\sum_{i\in A} w_i\hat{\pi}_i^{-1}R_i\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\sum_{i\in A} w_i\hat{\pi}_i^{-1}R_i\mathbf{x}_iy_i. \end{aligned}$$

Note that the estimator $\bar{y}_{re}$ is a linear function of y_i's in the respondents. That is, we can write $\bar{y}_{re} = \sum_{i\in A} w_i^*R_iy_i$ for some w_i^*. The new weight w_i^* can also be obtained as follows:

$$\min_{w_i^*}\sum_{i\in A} R_iw_i\hat{\pi}_i^{-1}\left(w_i^*-w_i\hat{\pi}_i^{-1}\right)^2$$

subject to

$$\sum_{i\in A} w_i\mathbf{x}_i = \sum_{i\in A} w_i^*R_i\mathbf{x}_i. \tag{11}$$

Thus, the regression weight w_i^* are constructed to satisfy the constraint (11). The regression NWA estimator using w_i^* as the final weight can improve the efficiency significantly if the study variable y_i is well approximated by a linear combination of the elements of $\mathbf{x}_i$.

Assume that $\mathbf{x}_i$ is q-dimensional and $\mathbf{x}_i = \left(1, \mathbf{x}_{1i}'\right)'$, expression (10) can be changed as

$$\bar{y}_{re} = \bar{y}_e + \left(\bar{\mathbf{x}}_{1n}-\bar{\mathbf{x}}_{1e}\right)'\hat{\boldsymbol{\beta}}_{1e}, \tag{12}$$

where $\bar{\mathbf{x}}_{1n}$ and $\bar{\mathbf{x}}_{1e}$ are the vectors of the last $q-1$ components of $\bar{\mathbf{x}}_n$ and $\bar{\mathbf{x}}_e$, respectively, and $\hat{\boldsymbol{\beta}}_{1e}$ is the vector of the last $q-1$ components of $\hat{\boldsymbol{\beta}}_e$. Then, under conditions such as those used by Isaki and Fuller (1982),

$$\begin{aligned} \bar{y}_{re} &= \bar{y}_e + \left(\hat{\mathbf{x}}_{1n}-\hat{\mathbf{x}}_{1e}\right)'\boldsymbol{\beta}_1 + O_p\left(n^{-1}\right), \\ &= \bar{y}_N + \bar{y}_e - \left(\boldsymbol{\beta}_0 + \hat{\mathbf{x}}_{1e}'\boldsymbol{\beta}_1\right) + \left(\hat{\mathbf{x}}_{1n}-\bar{\mathbf{x}}_N\right)'\boldsymbol{\beta}_1 \\ &\quad + O_p\left(n^{-1}\right) \end{aligned} \tag{13}$$

where $\boldsymbol{\beta}_1 = p\lim\hat{\boldsymbol{\beta}}_{1e}$, $\beta_0 = \bar{y}_N - \bar{\mathbf{x}}_N'\boldsymbol{\beta}_1$, and $\bar{\mathbf{x}}_N = N^{-1}\sum_{i=1}^N \mathbf{x}_i$. If we define $a_i = y_i - \beta_0 - \mathbf{x}_i'\boldsymbol{\beta}_1$ and use the similar argument for (8), $\bar{y}_r$ in (13) can be written

$$\begin{aligned} \bar{y}_{re} &= \hat{a}_d + \beta_0 + \hat{\mathbf{x}}_{1n}'\boldsymbol{\beta}_1 - \Delta_{az}'\left(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\right) + O_p\left(n^{-1}\right) \\ &= \bar{y}_{rd} - \Delta_{az}'\left(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\right) + O_p\left(n^{-1}\right), \end{aligned} \tag{14}$$

where $\hat{a}_d = \sum_{i\in A} w_i\pi_i^{-1}R_ia_i$, $\Delta_{az} = E\left\{\sum_{i\in A} w_i\left(1-\pi_i\right)a_iH\left(\mathbf{z}_i, \boldsymbol{\alpha}^0\right) \mid \mathcal{F}_N\right\}$, and $\bar{y}_{rd} = \bar{y}_d + \left(\hat{\mathbf{x}}_{1n}-\hat{\mathbf{x}}_{1d}\right)'\boldsymbol{\beta}_1$. Thus, similarly to (9), we have

$$\begin{aligned} Var\left(\bar{y}_{re} \mid \mathcal{F}_N\right) &= Var\left(\bar{y}_{rd} \mid \mathcal{F}_N\right) - \Delta_{az}'V_{22}\Delta_{az} \\ &\leq Var\left(\bar{y}_{rd} \mid \mathcal{F}_N\right). \end{aligned} \tag{15}$$

When the original sample is partitioned into G exhaustive and exclusive cells, the regression estimator in (10) with the auxiliary variable $\mathbf{x}_i = \left(x_{i1}, x_{i2}, \cdots, x_{iG}\right)$, where x_{ig} takes the value one if unit i belongs to cell g and takes zero otherwise, can be written

$$\bar{y}_{re} = \sum_{g=1}^G\left(\sum_{i\in A} w_ix_{ig}\right)\frac{\sum_{i\in A} w_i\hat{\pi}_i^{-1}R_ix_{ig}y_i}{\sum_{i\in A} w_i\hat{\pi}_i^{-1}R_ix_{ig}}.$$

Here, we do not have to assume the equal response probability within each cell because the nonresponse bias is already taken care of by the direct inclusion of the estimated response probability. The efficiency can be improved if the cells are formed with homogeneous y-values within each cell.

4. Variance estimation

We only discuss the variance estimation for the direct NWA estimator. Variance estimation for the regression NWA estimator can be derived similarly. From equation (9), a consistent estimator for the variance of $\bar{y}_e$ can be derived as

$$\hat{V}_e = \hat{V}_{11} - \hat{\Delta}'_{yz}\hat{V}_{22}\hat{\Delta}_{yz}, \tag{16}$$

where $\hat{V}_{11}$ is a consistent estimator for the variance of $\bar{y}_d$, $\hat{\Delta}_{yz} = \sum_{i\in A} w_i R_i\left(\hat{\pi}_i^{-1}-1\right)(y_i-\bar{y}_e) H(\mathbf{z}_i;\hat{\boldsymbol{\alpha}})$, and $\hat{V}_{22} = [I(\hat{\boldsymbol{\alpha}})]^{-1}$.

To derive a consistent estimator $\hat{V}_{11}$, let $\hat{V}_n$ be the full sample estimator of the variance of $\bar{y}_n$ of the form $\hat{V}_n = \sum_{i\in A}\sum_{j\in A}\Omega_{ij}y_iy_j$. If π_i were known, a consistent variance estimator of V_{11} can be obtained by

$$V_{11}^* = \sum_{i\in A}\sum_{j\in A}\Omega_{ij}\left\{\frac{R_i}{\pi_i}(y_i-\bar{y}_d)\right\}\left\{\frac{R_j}{\pi_j}(y_j-\bar{y}_d)\right\}. \tag{17}$$

As the above formula is a quadratic function of π_i^{-1}'s, by (7), using $\hat{\pi}_i$ instead of π_i in (17) is not consistent to the original V_{11}. From (7), we have

$$\pi_i^{-1} = \hat{\pi}_i^{-1} + \left(\hat{\pi}_i^{-1}-1\right)H\left(\mathbf{z}_i;\boldsymbol{\alpha}^0\right)'\left(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\right) + O_p\left(n^{-1}\right). \tag{18}$$

Thus, inserting (18) into (17), a consistent estimator of V_{11} can be derived as

$$\begin{aligned}\hat{V}_{11} &= \sum_{i\in A}\sum_{j\in A}\Omega_{ij}\left\{\frac{R_i}{\hat{\pi}_i}(y_i-\bar{y}_e)\right\}\left\{\frac{R_j}{\hat{\pi}_j}(y_j-\bar{y}_e)\right\}\\ &\quad + trace\left\{QI(\hat{\boldsymbol{\alpha}})^{-1}\right\}\end{aligned}$$

where

$$Q = \sum_{i\in A}\sum_{j\in A}\Omega_{ij}\left\{R_i(y_i-\bar{y}_e)\tilde{H}_i\right\}\left\{R_j(y_j-\bar{y}_e)\tilde{H}_j'\right\}$$

and $\tilde{H}_i = \left(\hat{\pi}_i^{-1}-1\right)H(\mathbf{z}_i;\hat{\boldsymbol{\alpha}})$.

5. Simulation Studies

To test our theory, we performed a limited simulation study. In the simulation study, $B = 5,000$ samples of size n were generated with

$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} \overset{i.i.d.}{\sim} N\left[\begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{xy} & 0 \\ \rho_{xy} & 1 & \rho_{zy} \\ 0 & \rho_{zy} & 1 \end{pmatrix}\right],$$

where $\overset{i.i.d.}{\sim}$ stands for independently and identically distributed. Two values of sample size, $n = 100$ and $n = 200$, and three levels of the population parameters, $(\rho_{xy}, \rho_{zy}) = (0.7, 0.7), (0.7, 0), (0, 0.7)$, were used in the simulation. The response indicator variable R_i were also generated from a Bernoulli distribution with parameter $\pi_i = [1 + exp(-Z_i)]^{-1}$. The parameter of interest is the population mean of Y and four estimators were computed from each sample. The four estimators are

1. $\bar{y}_d = \left(\sum_{i=1}^n \pi_i R_i\right)^{-1}\sum_{i=1}^n \pi_i R_i Y_i$: direct NWA estimator using the true response probability π_i.
2. $\bar{y}_e = \left(\sum_{i=1}^n \hat{\pi}_i R_i\right)^{-1}\sum_{i=1}^n \hat{\pi}_i R_i Y_i$: direct NWA estimator using the estimated response probability $\hat{\pi}_i$.
3. $\bar{y}_{rd}$: regression NWA estimator with the auxiliary variable $(1, X_i)$ using the true response probability π_i.
4. $\bar{y}_{rd}$: regression NWA estimator with the auxiliary variable $(1, X_i)$ using the estimated response probability $\hat{\pi}_i$.

In $\bar{y}_e$ and $\bar{y}_{re}$, the logistic regression coefficients are estimated iteratively by solving the following estimation equations using the Newton-Raphson method:

$$\sum_{i=1}^n (1, Z_i) R_i = \sum_{i=1}^n (1, Z_i)\hat{\pi}_i. \tag{19}$$

Also, the corresponding variance estimators for the point estimators were computed.

For each n and for each population parameters, the Monte Carlo mean and the Monte Carlo variance of the four point estimators are presented in Table 1. Monte Carlo variances in Table 1 show that the regression estimators are significantly more efficient than the direct estimators when $\rho_{xy} = 0.7$, which is consistent with the theory. Also, the estimators using the estimated response probability are significantly more efficient than the estimators using the true response probability when $\rho_{zy} = 0.7$. This is because the Δ_{zy} term in (9) is essentially estimating the covariance between Y_i and Z_i in the nonrespondents. Thus, larger value of ρ_{zy} leads to bigger reduction in the variance of the estimator using the estimated response probability. The amount of reduction of variance from $\bar{y}_d$ to $\bar{y}_e$ is about the same as the amount of variance from $\bar{y}_d$ to $\bar{y}_{rd}$ when the two correlations, ρ_{xy} and ρ_{zy}, are the same. The estimator $\bar{y}_{re}$, regression estimator using the estimated response probability, incorporates all the available information and thus shows the best efficiencies in most cases.

Table 2 presents the relative biases and the t-statistics of the variance estimators computed from the 5,000 Monte Carlo samples. The relative bias of the estimated variance, denoted by R.B. in the table, is the Monte Carlo bias divided by the Monte Carlo mean. The t-statistic for testing the hypothesis of zero bias is the Monte Carlo estimated bias divided by the Monte Carlo standard error of the estimated bias. The relative biases and the t-statistics are all near zero and decrease in absolute values as n increases.

The above simulation results show that the NWA estimators and the proposed estimators for the variances of NWA estimators are approximately unbiased and show good finite sample performances under the logistic regression model for the response probability. Adjustment using the estimated response probability improves the efficiency of the point estimator, in addition to the bias, because it incorporates additional information contained in the auxiliary variable used in the response model. Further gain in efficiency can also be achieved using a regression weighting method.

Table 1: Means and variances of the point estimators for each parameters, based on 5,000 samples.

n	(ρ_{xy}, ρ_{zy})	Estimator	Mean	Variance
100	$(0.7, 0)$	$\bar{y}_d$	2.00	0.0158
		$\bar{y}_e$	2.00	0.0161
		$\bar{y}_{rd}$	2.00	0.0133
		$\bar{y}_{re}$	2.00	0.0135
	$(0, 0.7)$	$\bar{y}_d$	2.00	0.0187
		$\bar{y}_e$	2.00	0.0161
		$\bar{y}_{rd}$	2.00	0.0185
		$\bar{y}_{re}$	2.00	0.0156
	$(0.7, 0.7)$	$\bar{y}_d$	2.00	0.0185
		$\bar{y}_e$	2.00	0.0149
		$\bar{y}_{rd}$	2.00	0.0156
		$\bar{y}_{re}$	2.00	0.0117
200	$(0.7, 0)$	$\bar{y}_d$	2.00	0.00797
		$\bar{y}_e$	2.00	0.00811
		$\bar{y}_{rd}$	2.00	0.00655
		$\bar{y}_{re}$	2.00	0.00662
	$(0, 0.7)$	$\bar{y}_d$	2.00	0.00932
		$\bar{y}_e$	2.00	0.00731
		$\bar{y}_{rd}$	2.00	0.00925
		$\bar{y}_{re}$	2.00	0.00722
	$(0.7, 0.7)$	$\bar{y}_d$	2.00	0.00954
		$\bar{y}_e$	2.00	0.00754
		$\bar{y}_{rd}$	2.00	0.00781
		$\bar{y}_{re}$	2.00	0.00577

Table 2: Relative biases and t-statistics of the variance estimators, based on 5,000 samples.

n	(ρ_{xy}, ρ_{zy})	Parameter	R. B. (%)	t-statistic
100	$(0.7, 0)$	$Var(\bar{y}_d)$	1.96	1.02
		$Var(\bar{y}_e)$	-0.78	-0.41
		$Var(\bar{y}_{rd})$	-0.95	-0.48
		$Var(\bar{y}_{re})$	-2.74	-1.39
	$(0, 0.7)$	$Var(\bar{y}_d)$	2.23	1.09
		$Var(\bar{y}_e)$	-12.72	-2.44
		$Var(\bar{y}_{rd})$	3.28	1.60
		$Var(\bar{y}_{re})$	-8.04	-1.91
	$(0.7, 0.7)$	$Var(\bar{y}_d)$	2.61	1.27
		$Var(\bar{y}_e)$	-4.78	-2.11
		$Var(\bar{y}_{rd})$	1.67	0.82
		$Var(\bar{y}_{re})$	-1.80	-0.82
200	$(0.7, 0)$	$Var(\bar{y}_d)$	0.83	0.41
		$Var(\bar{y}_e)$	-0.90	-0.44
		$Var(\bar{y}_{rd})$	0.15	0.08
		$Var(\bar{y}_{re})$	-0.85	-0.42
	$(0, 0.7)$	$Var(\bar{y}_d)$	2.07	1.00
		$Var(\bar{y}_e)$	-0.59	-0.28
		$Var(\bar{y}_{rd})$	2.48	1.19
		$Var(\bar{y}_{re})$	0.96	0.46
	$(0.7, 0.7)$	$Var(\bar{y}_d)$	-0.36	-0.18
		$Var(\bar{y}_e)$	-3.42	-1.68
		$Var(\bar{y}_{rd})$	1.47	0.75
		$Var(\bar{y}_{re})$	0.10	0.05

References

Bethlehem, J. G. (1988). Reduction of nonresponse bias through regression estimation, *Journal of Official Statistics*, **4**, 251-260.

Brick, J. M. and Kalton, G. (1996). Handling missing data in survey research. *Statist. Meth. Med. Res.* **5**, 215-238.

Ekholm, A. and Laaksonen, S. (1991). Weighting via response modeling in the Finnish household budget survey. *Journal of Official Statistics*, **7**, 325-337.

Eltinge, J. L. and Yansaneh, I. S. (1997). Diagnostics for formation of nonresponse adjustment cells, with an application to income nonresponse in the U.S. consumer expenditure survey. *Survey Methodology*, **23**, 33-40.

Isaki, C. and Fuller, W. A. (1982). Survey design under the regression superpopulation model.

Journal of the American Statistical Association, **77**, 89-96.

Fuller, W. A. (2002). Regression estimation for sample surveys. *Survey Methodology*, **28**, 5-23.

Fuller, W. A. and An, A. B. (1998). Regression adjustment for nonresponse. *Jour. Ind. Soc. Ag. Statistics*, **51**, 331-342.

Groves, R., Dillman, D., Eltinge, J., and Little, R.J.A. (2002). *Survey Nonresponse*. New York: Wiley.

Iannacchione, V. G. (2003). Sequential weight adjustment for location and cooperation propensity for the 1995 national survey of family growth, *Journal of Official Statistics*, **19**, 31-43.

Isaki, C.T. and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, **77**, 89-96.

Little, R.J.A. (1986). Survey nonresponse adjustments. *International Statistical Review*, **54**, 139-157.

Rizzo, L., Kalton, G., and Brick, J. M. (1996). A comparison of some weighting adjustment methods for panel nonresponse. *Survey Methodology*, **22**, 43-53.

Robins, J.M., Rotnitzky, A., and Zhao, L.P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, **89**, 846-866.

Rosenbaum, P.R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association*, **82**, 387-394.

Lundström, S. and Särndal, C.-E. (1999). Calibration as a standard method for treatment of nonresponse. *Journal of Official Statistics*, **15**, 305-327.