Nonresponse weighting adjustment using estimated response probability

Jae Kwang Kim and Jay J. Kim

Key words and phrases: Nonresponse; propensity score; regression estimator; survey sampling; unit nonresponse; variance estimation.

MSC 2000: Primary 62D05; secondary 62G09.

Abstract: To reduce nonresponse bias in sample surveys, a method of nonresponse weighting adjustment is often used which consists of multiplying the sampling weight of the respondent by the inverse of the estimated response probability. The authors examine the asymptotic properties of this estimator. They prove that it is generally more efficient than an estimator which uses the true response probability, provided that the parameters which govern this probability are estimated by maximum likelihood. The authors discuss variance estimation methods that account for the effect of using the estimated response probability; they compare their performances in a small simulation study. They also discuss extensions to the regression estimator.

Correction de la non-réponse par repondération au moyen d'une estimation de la probabilité de réponse

Résumé: Pour réduire le biais dû à la non-réponse dans les enquêtes, on fait souvent appel à une méthode d'ajustement dans laquelle le poids de sondage de chaque répondant est multiplié par l' inverse d'une estimation de la probabilité de réponse. Les auteurs étudient les propriétés asymptotiques de cet estimateur. Ils démontrent qu'il est généralement plus efficace que celui qui fait intervenir la probabilité de réponse théorique, pourvu que les paramètres qui régissent cette probabilité soient estimés par vraisemblance maximale. Les auteurs évoquent diverses méthodes d'estimation de la variance qui tiennent compte du fait que la probabilité de réponse est estimée; ils en comparent la performance dans le cadre d'une petite étude de simulation. Ils étendent aussi leurs résultats à l'estimateur obtenu par régression.

1. INTRODUCTION

Weighting adjustment is a popular method for handling unit nonresponse in sample surveys. Groves, Dillman, Eltinge & Little (2002) and Särndal & Lundström (2006) provided comprehensive overviews of nonresponse weighting adjustment (NWA) methods in survey sampling. In order for the respondents to properly represent the original population, the sampling weight of the respondent is increased using the information observed in the sample.

A theory for two-phase sampling, according to which the set of respondents is treated as a second phase sample from the original sample, leads us to multiply the inverse of the response probability by the sampling weight of the respondent. Since the true response probability is usually unknown, the estimated response probability is used to correct for nonresponse bias. When the estimated response probability is directly used and no other adjustment is made, the method is called the direct NWA method. See Rosenbaum (1987). Applications of the direct NWA method can be found in Ekholm & Laaksonen (1991), Folsom & Singh (2000), and Iannacchione (2003). The direct NWA estimator is designed to reduce nonresponse bias.

When auxiliary variables correlated with the study variable of interest are also available throughout the sample, we can improve efficiency of the direct NWA estimator by taking advantage of the auxiliary variables in the estimation procedure. Regression weighting is one of the commonly used techniques for incorporating the auxiliary variables in this situation. Cassel, Särndal & Wretman (1983), Bethlehem (1988), Fuller & An (1998), and Lundström & Särndal (1999) discuss the regression NWA method.
The NWA estimators discussed above reduce nonresponse bias because the estimated response probability is incorporated into the estimator. Although commonly used in practice, to the knowledge of the authors, asymptotic properties of the NWA estimators using the estimated response probability are not fully discussed in the literature. Some authors, including Rosenbaum (1987), Robins, Rotnitzky & Zhao (1994), and Little & Vartivarian (2005), 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. One notable exception is Beaumont (2005), who gave a clear justification for reduced variance using estimated response probability from a logistic regression model in the imputation context.

In this paper, we extend Beaumont’s results to more general response probability models and to a more general class of point estimators. We show that the estimator using the estimated response probability is more efficient than the estimator using the true response probability when the parameters for response probabilities are estimated by the maximum likelihood method. In Section 2, we provide details of a proof of consistency of the direct NWA estimator. In Section 3, variance estimation methods for the direct NWA estimator are discussed. In Section 4, results from a Monte Carlo study are presented. In Section 5, extension to the regression NWA estimator is discussed. Concluding remarks are given in Section 6.

2. MAIN RESULTS

Let the finite population be of size \( N \), indexed from 1 to \( N \), where \( N \) is assumed 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 variable 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
\[
\bar{y}_n = N^{-1} \sum_{i \in A} \pi_i^{-1} y_i,
\]
where \( \pi_i = \Pr(i \in A) \) and \( A \) is the set of indices in the sample. The weight \( \pi_i^{-1} \) is called the sampling weight of unit \( i \). By the definition of \( \pi_i \), we have
\[
E(\bar{y}_n | F_N) = \bar{y}_N, \tag{1}
\]
where \( F_N = \{ u_1, \ldots, u_N \} \), \( u_i = (x_i^T, y_i) \), and \( x_i \) is the vector of auxiliary variables for unit \( i \). The expectation (1) is taken with respect to the sampling mechanism, which is generated by repeated application of the prescribed sample selection method.

Under nonresponse, we define the response indicator variable of \( y_i \) by
\[
R_i = \begin{cases} 
1 & \text{unit } i \text{ responds,} \\
0 & \text{unit } i \text{ does not respond,}
\end{cases}
\]
for \( i \in A \), and let \( p_i|A = \Pr(R_i = 1 | i \in A) \) be the response probability of sampled unit \( i \). If we know the response probability \( p_i|A \), then the population mean can be unbiasedly estimated by multiplying the sampling weight by the inverse of \( p_i|A \). Let
\[
\bar{y}_d = \frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \frac{R_i}{p_i|A} y_i \tag{2}
\]
be such an estimator constructed by directly using the true response probability. By definition of \( p_i|A \), the conditional expectation of \( \bar{y}_d \), conditional on \( A \), is equal to \( \bar{y}_n \) and, hence, \( \bar{y}_d \) is unbiased for \( \bar{y}_N \).

When the true response probability \( p_i|A \) is not available, we use its estimate \( \hat{p}_i|A \) obtained from the model for the response probability. Let
\[
\bar{y}_e = \frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \frac{R_i}{\hat{p}_i|A} y_i \tag{3}
\]
be an estimator of the population mean using the estimated response probability. The estimator (3) is called the direct NWA estimator because an estimated response probability is used and no auxiliary variables are incorporated into the estimator. The estimator \( \hat{y}_{ij} \) in (2) is not applicable in practice because the true response probability is not known in general.

To formally discuss asymptotic properties of the direct NWA estimator, we assume a sequence of samples and finite populations described in Isaki & Fuller (1982). Assume that the sequence of finite populations of \( u_i = (1, x_i^\top, y_i)^\top \) has bounded fourth moments. Assume that sample moments converge to population moments, viz

\[
N^{-1} \sum_{i=1}^{N} \pi_i^{-1} w_i \tilde{w}_i^\top - N^{-1} \sum_{i=1}^{N} w_i \tilde{w}_i^\top = O_P(n^{-1/2}),
\]

where \( w_i = \text{vec}(u_i u_i^\top) \) is the column vector obtained by stacking the columns of the matrix \( u_i u_i^\top \). Here, it is understood that \( N \) is indexed by \( n \) as \( N_n \), but the subscript \( n \) is omitted for brevity. We assume that no extreme weights dominate the others, in that

\[
K_{\pi,1} \leq n^{-1} N \pi_i \leq K_{\pi,2}
\]

for all \( i = 1, \ldots, N \), uniformly in \( n \), where \( K_{\pi,1} \) and \( K_{\pi,2} \) are fixed constants.

For the response mechanism, we assume the following conditions.

[R.1] The responses are independent,

\[
\text{cov}(R_i, R_j \mid A, \mathcal{F}) = \begin{cases} 
    p_{iA}(1 - p_{iA}) & \text{if } i = j, \\
    0 & \text{otherwise.}
\end{cases}
\]

[R.2] The response probability is parametrically modeled; that is, the probability of response is

\[
p_{iA} = p(z_i; \alpha_A^0),
\]

for some known function \( p(z_i; \cdot) \) with parameter \( \alpha \) evaluated at \( \alpha = \alpha_A^0 \), where \( z_i \) is a vector of variables observed for both respondents and nonrespondents. We assume that \( p(z_i; \alpha) \) is continuous in \( \alpha \) with continuous first and second derivatives in an open set containing \( \alpha_A^0 \) for all \( z_i \) and for all possible samples \( A \).

[R.3] The response probability is bounded below; that is, \( p_{iA}^{-1} < K_p \) for all \( i \in A \) and for all possible samples \( A \), uniformly in \( n \), where \( K_p \) is a fixed constant.

We also assume that the true value \( \alpha_A^0 \) is estimated by \( \hat{\alpha} \), which is the unique solution to the estimating equation

\[
\frac{\partial}{\partial \alpha} \sum_{i \in A} k_i \{ R_i \ln(p_{iA}) + (1 - R_i) \ln(1 - p_{iA}) \} = 0,
\]

where \( k_i \) is the weight of unit \( i \) in the estimating equation for \( \alpha \). When \( k_i = 1 \), the solution to (7) is the usual maximum likelihood estimate for \( \alpha_A^0 \). Beaumont (2005) advocated using \( k_i = 1 \) under the logistic regression model for the response probability. Under the two-phase sampling approach, \( k_i = \pi_i^{-1} \) is also commonly used. For example, see Fuller & An (1998).

Once \( \hat{\alpha} \) is computed, the true response probability \( p_{iA} \) is estimated by \( \hat{p}_{iA} = p(z_i; \hat{\alpha}) \). If we define \( h_i(\alpha) = \partial \ln(p_{iA}) / \partial \alpha \) with \( \logit(p) = \ln(p/(1 - p)) \), the estimating equation (7) can be written

\[
S(\alpha) = \sum_{i \in A} k_i \{ R_i - p(z_i; \alpha) \} h_i(\alpha) = 0.
\]
Note that $\alpha^0_A$ satisfies
\[ \mathbb{E} \{ S(\alpha^0_A) \mid A, \mathcal{F}_N \} = 0, \] (9)
regardless of the choice of $k_i$ in (8), where the expectation in (9) is over the response mechanism, conditional on the sample.

Also, we assume that
\[ p \lim_{n \to \infty} N^{-1} \sum_{i \in A} \frac{1}{\pi_i} \left[ h_i(\alpha), h_i(\alpha)h_i(\alpha)^\top, \{ \partial h_i(\alpha)/\partial \alpha \} \right] y_i < \infty \] (10)
uniformly in $\alpha$. For example, for the logistic regression model where $\text{logit}(p_{iA}) = z_i^\top \alpha$, we have $h_i(\alpha) = z_i$ and $h_i(\alpha)$ satisfies (10) when $z_i$ has finite second moments.

Under the regularity conditions, conditional on the sample, the estimator $\hat{\alpha}$ satisfies
\[ \hat{\alpha} - \alpha^0_A = -\{ I(\alpha^0_A) \}^{-1} S(\alpha^0_A) + o_p(n^{-1/2}), \] (11)
where $I(\alpha^0_A)$ is the expected information matrix defined by
\[ I(\alpha^0_A) = -\mathbb{E} \left\{ \frac{\partial}{\partial \alpha} S(\alpha^0_A) \mid A, \mathcal{F}_N \right\}. \]

The relationship in (11) essentially follows from (9). The reference distribution for (11) is the response mechanism, conditional on the sample. Thus, the asymptotic equivalence of $\hat{\alpha}^\ast$ to $\bar{y}_e$ also holds unconditionally. The linearized term, $\bar{y}_{el}$, satisfies
\[ \mathbb{E} (\bar{y}_{el} \mid \mathcal{F}_N) = \bar{y}_N \]
and
\[ V(\bar{y}_{el} \mid \mathcal{F}_N) = V(\bar{y}_n \mid \mathcal{F}_N) + \frac{1}{N^2} \mathbb{E} \left\{ \sum_{i \in A} \frac{1}{\pi_i^2} \frac{1 - p_{iA}}{p_{iA}} (y_i - k_i \pi_i p_{iA} h_{i0}^\top \gamma_n)^2 \mid \mathcal{F}_N \right\}. \] (14)
Theorem 1 states that the direct NWA estimator is asymptotically equivalent to a random variable that is unbiased for the population mean. To discuss the variance, consider a class of estimators of the form

\[
\hat{y}_\gamma = \frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \left\{ k_i \pi_i p_{i|A} h_{i0}^T \gamma + \frac{R_i}{p_{i|A}} (y_i - k_i \pi_i p_{i|A} h_{i0}^T \gamma) \right\},
\]

which is indexed by \( \gamma \). If we choose \( \gamma = 0 \), then \( \hat{y}_\gamma \) reduces to \( \hat{y}_d \), the direct NWA estimator using the true response probability. Note that \( \hat{y}_\gamma \) is unbiased for \( \bar{y}_N \) for any choice of \( \gamma \) and its variance is

\[
V(\hat{y}_\gamma | \mathcal{F}_N) = \frac{1}{N^2} \mathbb{E} \left\{ \sum_{i \in A} \frac{1}{\pi_i} \frac{1 - p_{i|A}}{p_{i|A}} (y_i - k_i \pi_i p_{i|A} h_{i0}^T \gamma)^2 \right\},
\]

which takes its minimum at \( \gamma = \gamma^* \), where

\[
\gamma^* = \left\{ \sum_{i \in A} k_i^2 p_{i|A} (1 - p_{i|A}) h_{i0}^T \right\}^{-1} \sum_{i \in A} \pi_i^{-1} k_i (1 - p_{i|A}) h_{i0} y_i.
\]

Thus, when \( k_i = 1 \), we have \( \gamma^* = \gamma_n \) where \( \gamma_n \) is defined in (13). By (12), the estimator \( \hat{y}_\gamma \) using \( \gamma = \gamma_n \) is asymptotically equivalent to the direct NWA estimator \( \hat{y}_e \). Therefore, the direct NWA estimator \( \hat{y}_d \) using the estimated response probability computed from the estimating equation (7) with \( k_i = 1 \) is asymptotically equivalent to a random variable, which is no less efficient than the NWA estimator \( \hat{y}_d \) using the true response probability. That is, for \( k_i = 1 \),

\[
V(\hat{y}_e | \mathcal{F}_N) \leq V(\hat{y}_d | \mathcal{F}_N).
\]

Result (17) shows that efficiency of \( \hat{y}_d \) can be improved if we use the estimated response probability computed from the maximum likelihood method. In Beaumont (2005), an efficiency gain was claimed for \( k_i = 1 \), the maximum likelihood equation for \( \alpha \), under the logistic regression model for the response probability. Here, the result is proved under a more general class of the response model.

**Remark 1.** Theorem 1 is derived for general \( k_i \) in (7). Thus, consistency of the direct NWA estimator holds regardless of the choice of \( k_i \). However, efficiency can be different for different choices of \( k_i \). Note that, if we treat \( \hat{y}_\gamma \) in (15) as a class of estimators indexed by \( k_i \), variance in (16) is minimized for

\[
k_i = \frac{1}{\pi_i} \frac{1}{p_{i|A}} \frac{y_i}{h_{i0}^T \gamma}
\]

for some \( \gamma \). Thus, although the choice of \( k_i = 1 \) may lead to the best estimate of \( \alpha \) and the best estimate of the individual response probability, it does not necessarily provide the best NWA estimates because the variance of \( \hat{y}_e = \hat{y}_e(\alpha_k) \) with \( k_i = \pi_i^{-1} \) can be smaller than the variance of \( \hat{y}_e = \hat{y}_e(\alpha_k) \) with \( k_i = 1 \). That is, the maximum likelihood method is not necessarily efficient for the NWA purpose. Further discussion on the optimal NWA estimation is beyond the scope of the paper and will be a topic of future research.

### 3. Variance Estimation

We now discuss the issue of variance estimation for the NWA estimators. Under complete response, we consider variance estimators of the form

\[
\hat{V}_n(\bar{y}_n) = \sum_{i \in A} \sum_{j \in A} \Omega_{ij} y_i y_j
\]
for some coefficients $\Omega_{ij}$. We assume that the variance estimator $\hat{V}_n(\hat{y}_n)$ is unbiased for the variance of $\hat{y}_n$.

To derive a variance estimator of the direct NWA estimator, we first consider the reverse approach of Fay (1991) and Shao & Steel (1999). In the reverse approach, the definition of $R_i$ is extended to the entire finite population, as $R_i = 1$ if unit $i$ responds when sampled and $R_i = 0$ if unit $i$ does not respond when sampled. To apply the reverse approach properly, we assume that the response probability of a unit does not depend on the characteristics of the other elements in the sample. Thus, we can write $p_i|A = p_i$.

Now, the linearized term $\bar{y}_{el}$ in (12) can be written

$$\bar{y}_{el} = \frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \eta_i,$$

where $\eta_i = k_i \pi_i \hat{p}_i \hat{h}_i^\top \gamma_N + R_i (y_i - k_i \pi_i \hat{p}_i \hat{h}_i^\top \gamma_N) / p_i$. The total variance of $\bar{y}_{el}$ can be written

$$V(\bar{y}_{el} | F_N) = E\{V(\bar{y}_{el} | R_N, F_N) | F_N\} + V\{E(\bar{y}_{el} | R_N, F_N) | F_N\},$$

(18)

where $R_N = (R_1, \ldots, R_N)$. Here, the conditional distribution given $R_N$ is the distribution over the sampling mechanism treating the response indicator $R_i$ as fixed. The conditional variance in (18) can be written

$$V(\bar{y}_{el} | R_N, F_N) = V\left(\sum_{i \in A} \frac{1}{\pi_i} \eta_i \bigg| R_N, F_N\right) = E\left(\sum_{i \in A} \sum_{j \in A} \Omega_{ij} \eta_i \eta_j \bigg| R_N, F_N\right).$$

Since we do not know $p_i$ and $h_{i0}$ in $\eta_i$, we use their estimates $\hat{p}_i = p(z_i; \hat{\alpha})$ and $\hat{h}_i = h(z_i; \hat{\alpha})$ to get a plug-in estimator for the first component of (18) as

$$\hat{V}_{e1} = \sum_{i \in A} \sum_{j \in A} \Omega_{ij} \hat{\eta}_i \hat{\eta}_j,$$

(19)

where

$$\hat{\eta}_i = k_i \pi_i \hat{p}_i \hat{h}_i^\top \hat{\gamma}_N + \frac{R_i}{\hat{p}_i} (y_i - k_i \pi_i \hat{p}_i \hat{h}_i^\top \hat{\gamma}_N),$$

$$\hat{\gamma}_N = \left\{ \sum_{i \in A_N} k_i (1 - \hat{p}_i) \hat{h}_i \hat{h}_i^\top \right\}^{-1} \sum_{i \in A_N} \pi_i^{-1} (\hat{p}_i - 1) \hat{h}_i y_i,$$

and $A_N = \{ i \in A \; ; R_i = 1 \}$ is the set of respondents in the sample. Note that the variance estimator $\hat{V}_{e1}$ is approximately unbiased conditionally for the conditional variance $V(\bar{y}_{el} | R_N, F_N)$, conditional on $R_N$, and thus is also approximately unbiased for $E\{V(\bar{y}_{el} | R_N, F_N) | F_N\}$.

To estimate the second component of the total variance in (18), note that

$$V\{E(\bar{y}_{el} | R_N, F_N) | F_N\} = N^{-2} \sum_{i=1}^N (\hat{p}_i^{-1} - 1) (y_i - k_i \pi_i \hat{p}_i \hat{h}_i^\top \gamma_N)^2,$$

and so an approximately unbiased estimator for the second variance component is

$$\hat{V}_{e2} = \frac{1}{N^2} \sum_{i \in A_N} \pi_i^{-1} \hat{p}_i^{-2} (1 - \hat{p}_i) (y_i - k_i \pi_i \hat{p}_i \hat{h}_i^\top \hat{\gamma}_N)^2.$$

(20)

Thus, the variance estimator using the reverse approach is

$$\hat{V}_e = \hat{V}_{e1} + \hat{V}_{e2},$$

(21)
where \( \hat{V}_{c1} \) and \( \hat{V}_{c2} \) are defined in (19) and (20), respectively. When the sampling fraction \( N^{-1}n \) is negligible, \( \hat{V}_{c2} \) is of smaller order than \( \hat{V}_{c1} \) and thus can be safely ignored. The proposed variance estimator is very similar in spirit to that of Haziza & Rao (2006).

Instead of (21), another variance estimator can also be derived using the variance formula (14). Note that the total variance (14) can be written

\[
V(\bar{y}_n | \mathcal{F}_N) = V_{\text{sam}} + V_{\text{res}},
\]

where \( V_{\text{sam}} = \text{var}(\bar{y}_n | \mathcal{F}_N) \) and

\[
V_{\text{res}} = \frac{1}{N^2} E \left\{ \sum_{i \in A} \Omega_{ii} y_i^2 + \sum_{i \neq j \in A} \Omega_{ij} y_i y_j \mid \mathcal{F}_N \right\}.
\]

Note that \( V_{\text{sam}} \) can be written

\[
V_{\text{sam}} = E \left( \sum_{i \in A} \Omega_{ii} y_i^2 + \sum_{i \neq j \in A} \Omega_{ij} y_i y_j \mid \mathcal{F}_N \right).
\]

If the \( p_i \) are known, the first term in the right-hand side of (23) is unbiasedly estimated by \( \sum_{i \in A_R} \Omega_{ii} \hat{p}_i^{-1} y_i^2 \) and the second term is unbiasedly estimated by \( \sum_{i \neq j} \sum_{i, j \in A_R} \Omega_{ij} \hat{p}_i^{-1} \hat{p}_j^{-1} y_i y_j \). Thus, using \( \hat{p}_i \) instead of \( p_i \), an estimator for the sampling variance of \( \bar{y}_n \) is

\[
\hat{V}_{\text{sam}} = \sum_{i \in A_R} \Omega_{ii} \hat{p}_i^{-1} y_i^2 + \sum_{i \neq j} \sum_{i, j \in A_R} \Omega_{ij} \hat{p}_i^{-1} \hat{p}_j^{-1} y_i y_j.
\]

An approximately unbiased estimator for \( V_{\text{res}} \) is

\[
\hat{V}_{\text{res}} = \frac{1}{N^2} \sum_{i \in A_R} \frac{1 - \hat{p}_i}{\hat{p}_i^2} (y_i - k_i \pi_i \hat{p}_i \hat{h}_i^\top \hat{y}_N)^2.
\]

The alternative variance estimator using (24) and (25) is very similar in spirit to the variance estimator proposed by Beaumont (2005) under an imputation context. The alternative variance estimator can be called a two-phase variance estimator because the decomposition in (22) resembles the standard variance decomposition in the two-phase sampling.

The two variance estimators discussed in this section are all approximately unbiased and should behave similarly in large samples. In the following section, we compare the finite sample performances of the two variance estimators using a simulation study.

### 4. SIMULATION STUDY

A numerical experiment was performed to compare the estimators. In the experiment, three artificially stratified finite populations are generated from a multivariate normal distribution

\[
\begin{pmatrix}
y_{hi} \\
z_{hi}
\end{pmatrix} \overset{i.i.d.}{\sim} N \left( \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \quad h = 1, 2, 3, 4; \ i = 1, \ldots, N_h,
\]

where \( i.i.d. \) stands for independently and identically distributed and \( N_h = 1000, 2000, 3000, 4000 \) for \( h = 1, 2, 3, 4 \), respectively. Three levels (0, 0.3, 0.6) of \( \rho \) are used to generate the three populations.

From each of the realized finite populations, two sets of independent stratified random samples of size \( n = 100 \) and \( n = 400 \) are generated without replacement and the sample sizes are all equal \( (n_h = n/4, \ h = 1, 2, 3, 4) \) in each stratum. We also generated response indicator variable \( R_{hi} \) from the Bernoulli distribution with probability \( p_{hi} = 1/(1 + \exp(-z_{hi} + 1)) \). The
finite populations of \((y_{hi}, z_{hi}, R_{hi})\) are fixed in the Monte Carlo sampling. The study variable \(y_{hi}\) is observed if and only if \(R_{hi} = 1\). The auxiliary variable \(z_{hi}\) is observed throughout the sample. The Monte Carlo sample sizes are all \(B = 10,000\) and the average response rate is about 0.69 in the simulation.

Using the Monte Carlo samples generated above, the means and variances of the three NWA estimators are computed. The estimators are:

1. \(\bar{y}_d\): direct NWA estimator using the true response probability,
2. \(\bar{y}_e(1)\): direct NWA estimator using the estimated response probability with \(k_i = 1\) in (8),
3. \(\bar{y}_e(\pi^{-1})\): direct NWA estimator using the estimated response probability with \(k_i = \pi_i^{-1}\) in (8).

In computing the estimated response probability, the predictor variable \((1, z_i)\) is used in (8) and the estimating equation (8) is computed iteratively using the Newton–Raphson method.

| Table 1: Monte Carlo relative biases and variances of the NWA estimators, based on 10,000 samples. |
|---|---|---|---|---|---|---|
| \(n\) | Estimator | Relative Bias (%) | Variance |
| | | \(\rho = 0.0\) | \(\rho = 0.3\) | \(\rho = 0.6\) | \(\rho = 0.0\) | \(\rho = 0.3\) | \(\rho = 0.6\) |
| 100 | \(\bar{y}_d\) | -0.01 | 0.01 | -0.011 | 0.04825 | 0.04105 | 0.03306 |
| | \(\bar{y}_e(1)\) | 0.011 | -0.003 | 0.001 | 0.02771 | 0.02382 | 0.02053 |
| | \(\bar{y}_e(\pi^{-1})\) | 0.008 | -0.001 | 0.002 | 0.02313 | 0.02018 | 0.01737 |
| 400 | \(\bar{y}_d\) | -0.007 | -0.004 | 0.008 | 0.01214 | 0.01002 | 0.00826 |
| | \(\bar{y}_e(1)\) | -0.006 | 0.002 | 0.009 | 0.00630 | 0.00568 | 0.00476 |
| | \(\bar{y}_e(\pi^{-1})\) | -0.001 | 0.006 | 0.012 | 0.00511 | 0.00471 | 0.00406 |

Table 1 presents the Monte Carlo relative biases and variances of the NWA estimators obtained from the simulation study. The Monte Carlo relative bias is computed by the Monte Carlo standard error. Table 1 reveals that the relative biases of the three estimators are all small with absolute values less than 2%.

The variance results in Table 1 show that the variance of the NWA estimator using the estimated response probability is smaller than the variance of the NWA estimator using the true response probability, regardless of the correlation coefficient between \(y\) and \(z\). Note that the conditional variance of \(\bar{y}_d\) can be written \(N^{-2} \sum_{i \in A} \pi_i^{-2} (\pi_i^{-1} - 1) y_i^2\). Given the same number of respondents (i.e., \(\sum_{i \in A} R_i\) is fixed), this conditional variance is minimized if \(p_i \propto \pi_i y_i\), which is more likely to happen for higher value of \(\rho\) when \(p_i\) is an increasing function of \(z_i\). Thus, this conditional variance gets smaller for larger values of \(\rho\), which is consistent with the numerical results in Table 1. In Table 1, the NWA estimator using \(k_i = 1\) is less efficient than the NWA estimator using \(k_i = \pi_i^{-1}\), which is consistent with Remark 1.

In addition to the point estimators, we also computed two variance estimators for each of the \(\bar{y}_e\). One is the variance estimator using the reverse approach, as defined in (21), and the other is the two-phase variance estimator, as the sum of (24) and (25). Table 2 presents the percent relative biases and the t-statistics of the variance estimators computed from the 10,000 Monte Carlo samples. The relative bias of the estimated variance is the Monte Carlo bias divided by the Monte Carlo variance of the point estimator. 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. For \(n = 100\), we see negative relative bias values implying that the two variance estimators generally underestimate the true variance.
the fact that the second order term in the Taylor expansion, which is nonnegative by (A.1), is not
taken into account in the variance estimation. The bias of the variance estimator is essentially of
order \( n^{-2} \) and can be negligible for large sample sizes. Thus, the relative biases of the variance
estimators get substantially smaller as \( n \) increases from \( n = 100 \) to \( n = 400 \). For the sample
size of \( n = 100 \), the absolute values of the relative biases are smaller for larger \( \rho \) values because
the magnitude of the second order term gets smaller as \( \rho \) gets larger. In the simulation, the
two variance estimators show similar performances for moderate sample size. The variance
estimator using the reverse approach has some computational advantage when the sampling rate
is negligible.

**Table 2: Percent relative biases and \( t \)-statistics of the variance estimators, based on 10,000 samples.**

| \( n \) | Parameter Method | Relative Bias (%) \( \rho = 0.0 \) \( \rho = 0.3 \) \( \rho = 0.6 \) \( t \)-statistic \( \rho = 0.0 \) \( \rho = 0.3 \) \( \rho = 0.6 \) |
|---|---|---|---|---|---|---|---|
| 100 | Variance Reverse | -5.74 | -3.99 | -3.87 | -3.82 | -2.77 | -2.66 |
| | of \( \bar{y}_e(1) \) Two-phase | -6.35 | -4.36 | -4.11 | -4.16 | -2.98 | -2.78 |
| | of \( \bar{y}_e(\pi^{-1}) \) Two-phase | -4.90 | -4.41 | -2.62 | -2.91 | -3.02 | -1.83 |
| 400 | Variance Reverse | 0.15 | -1.40 | 1.27 | 0.10 | -0.99 | 0.91 |
| | of \( \bar{y}_e(1) \) Two-phase | 0.05 | -1.48 | 1.26 | 0.03 | -1.05 | 0.90 |
| | Variance Reverse | -0.12 | -0.81 | 0.75 | -0.08 | -0.58 | 0.54 |
| | of \( \bar{y}_e(\pi^{-1}) \) Two-phase | 0.41 | -0.35 | 1.27 | 0.29 | -0.25 | 0.90 |

We also computed interval estimators for 95% nominal coverage. Table 3 displays mean
lengths and actual coverages of 95% confidence intervals using the two types of variance estimators. The confidence intervals are \( (\hat{\theta} - 1.96\sqrt{\hat{V}}, \hat{\theta} + 1.96\sqrt{\hat{V}}) \), where \( \hat{\theta} \) is a point estimator and \( \hat{V} \) is its estimated variance. The confidence intervals based on weighted NWA estimator (\( k_i = \pi - 1 \)) show better performance in terms of mean length than the confidence intervals based on the unweighted NWA estimator (\( k_i = 1 \)) because the weighted NWA estimator is more
efficient than the unweighted NWA estimator in this simulation. As in Table 2, there are no sig-
nificant differences between the interval estimator computed from the reverse method and that
computed from the two-phase method.

**Table 3: Mean lengths and coverages of 95% confidence interval estimators, based on 10,000 samples.**

| \( n \) | Point Estimator Method | Mean Length \( \rho = 0.0 \) \( \rho = 0.3 \) \( \rho = 0.6 \) Coverage (%) \( \rho = 0.0 \) \( \rho = 0.3 \) \( \rho = 0.6 \) |
|---|---|---|---|---|---|---|---|
| 100 | \( \bar{y}_e(1) \) Reverse | 0.620 | 0.585 | 0.547 | 93.7 | 93.8 | 94.2 |
| | Two-phase | 0.618 | 0.584 | 0.545 | 93.2 | 93.5 | 93.8 |
| | \( \bar{y}_e(\pi^{-1}) \) Reverse | 0.560 | 0.534 | 0.503 | 93.8 | 93.6 | 94.2 |
| | Two-phase | 0.563 | 0.537 | 0.505 | 93.8 | 93.7 | 94.4 |
| 400 | \( \bar{y}_e(1) \) Reverse | 0.310 | 0.292 | 0.271 | 94.7 | 94.7 | 95.3 |
| | Two-phase | 0.309 | 0.292 | 0.271 | 94.7 | 94.7 | 95.3 |
| | \( \bar{y}_e(\pi^{-1}) \) Reverse | 0.279 | 0.267 | 0.250 | 94.7 | 94.9 | 95.2 |
| | Two-phase | 0.279 | 0.268 | 0.251 | 94.7 | 95.0 | 95.2 |
5. EXTENSION TO REGRESSION ESTIMATION

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

$$\bar{y}_{re} = \bar{y}_c + (\bar{x}_n - \bar{x}_e)^\top \hat{\beta}_e,$$

where

$$\bar{x}_n = N^{-1} \sum_{i \in A} \pi_i^{-1} x_i,$$

$$\hat{\beta}_e = \left( \sum_{i \in A} \pi_i^{-1} \hat{p}_i^{-1} R_i x_i x_i^\top \right)^{-1} \sum_{i \in A} \pi_i^{-1} \hat{p}_i^{-1} R_i x_i y_i$$

and $\bar{x}_e = N^{-1} \sum_{i \in A} \pi_i^{-1} \hat{p}_i^{-1} R_i x_i$. The regression NWA estimator can improve efficiency of the direct NWA estimator significantly if the study variable $y_i$ is well approximated by a linear combination of the elements of $x_i$.

Under the assumptions for (12), it can be shown that $\hat{\beta}_e - \beta_N = O_p(n^{-1/2})$ and the regression NWA estimator in (26) satisfies

$$\bar{y}_{re} = \bar{y}_c + (\bar{x}_n - \bar{x}_e)^\top \beta_N + O_p(n^{-1}),$$

where

$$\beta_N = \left( \sum_{i = 1}^N x_i x_i^\top \right)^{-1} \sum_{i = 1}^N x_i y_i.$$

Thus, result (12) can be directly applied to $\bar{y}_c$ and $\bar{x}_e$ in (27) to obtain

$$\bar{y}_{re} = N^{-1} \sum_{i \in A} \frac{1}{\pi_i} \left\{ x_i^\top \beta_N + k_i \pi_i p_i h_i^\top \alpha_N + \frac{R_i}{p_i} (y_i - x_i^\top \beta_N - k_i \pi_i p_i h_i^\top \alpha_N) \right\} + O_p(n^{-1})$$

(28)

where

$$\alpha_N = \left\{ \sum_{i = 1}^N k_i \pi_i p_i (1 - p_i) h_i^\top h_i \right\}^{-1} \sum_{i = 1}^N (1 - p_i) h_i (y_i - x_i^\top \beta_N).$$

(29)

Thus, under some regularity conditions, the regression NWA estimator is approximately unbiased and the variance is, ignoring the smaller order terms,

$$V(\bar{y}_{re} | \mathcal{F}_N) = V(\bar{y}_n | \mathcal{F}_N) + N^{-2} E \left\{ \sum_{i \in A} \frac{1}{\pi_i^2} \frac{1 - p_i}{p_i} (y_i - x_i^\top \beta_N - k_i \pi_i p_i h_i^\top \alpha_N)^2 \bigg| \mathcal{F}_N \right\}.$$

(30)

Given $\beta_N$, the choice of $\alpha_N$ in (29) minimizes the conditional variance of the regression NWA estimator. Thus, similarly to (17), efficiency of the regression NWA estimator can be improved by using the estimated response probability computed from the maximum likelihood equation (7).

If the auxiliary variable $x_i$ is observed throughout the population, the regression NWA estimator can be written

$$\bar{y}_{re2} = \bar{y}_c + (\bar{x}_N - \bar{x}_e)^\top \hat{\beta}_e,$$

(31)
where \( \bar{x}_N = N^{-1} \sum_{i=1}^N x_i \). Similarly to (28), the regression NWA estimator can be linearized to obtain

\[
\bar{y}_{re2} = \bar{x}_N^\top \beta_N + \frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \left\{ k_i \pi_i p_i h_i^\top \alpha_N + \frac{R_i}{p_i} (y_i - x_i^\top \beta_N - k_i \pi_i p_i h_i^\top \alpha_N) \right\} + O_p(n^{-1}).
\]

From the linearization (32), the regression NWA estimator (31) is also approximately unbiased and the approximate variance is

\[
V(\bar{y}_{re2} | \mathcal{F}_N) = N^{-2} V \left\{ \sum_{i \in A} \frac{1}{\pi_i} (y_i - x_i^\top \beta_n) \right\} | \mathcal{F}_N \] 
\[
+ N^{-2} E \left\{ \sum_{i \in A} \frac{1}{\pi_i} \frac{1 - p_i}{p_i} (y_i - x_i^\top \beta_n - k_i \pi_i p_i h_i^\top \alpha_n)^2 \right\} | \mathcal{F}_N \}.
\]

For variance estimation, two methods discussed in Section 3 can be used similarly. To derive the variance estimator using the reverse approach, the Taylor expansion (28) or (32) can be used to obtain a suitable \( \hat{\eta}_i \) for variance estimation. The variance estimators under two-phase approach can also be derived from the variance formula (30) or (33). In Beaumont (2005), the two-phase variance estimator using the decomposition (33) was proposed.

Remark 2. The direct NWA estimator in (3) is less appealing for practical use because the weights no longer sum to 1 and the estimator can be very unstable when \( \hat{p}_i \) is close to 0. A more useful estimator is

\[
\bar{y}_{re2} = \frac{\sum_{i \in A} \pi_i^{-1} \hat{p}_i^{-1} R_i y_i}{\sum_{i \in A} \pi_i^{-1} \hat{p}_i^{-1} R_i},
\]

which can be called the Hájek estimator applied to the NWA estimation. Note that the Hájek estimator is algebraically equivalent to the regression NWA estimator \( \bar{y}_{re2} \) in (31) with \( x_i = 1 \). Thus, asymptotic properties and variance estimators of the Hájek estimator can be derived using the theory for the regression NWA estimator.

6. CONCLUDING REMARKS

We have discussed asymptotic properties of the NWA estimators when the response probability is estimated using a parametric maximum likelihood method. Adjustment using the estimated response probability improves efficiency of the point estimator, in addition to reducing bias, because it incorporates additional information contained in the auxiliary variables used in the response model. In this case, the variance estimators discussed here account for the variance reduction due to estimating the response probability.

Further we can gain efficiency also by using a regression weighting method discussed in Section 5. As a final illustration, consider a situation when the original sample is partitioned into \( G \) exhaustive and exclusive cells. In this case, a commonly used NWA estimator is

\[
\bar{y}_r = N^{-1} \sum_{g=1}^G \left( \sum_{i \in A} \pi_i^{-1} x_{ig} \right) \frac{\sum_{i \in A} \pi_i^{-1} R_i x_{ig} y_i}{\sum_{i \in A} \pi_i^{-1} R_i x_{ig}},
\]

where \( x_{ig} \) takes the value one if unit \( i \) belongs to cell \( g \) and takes zero otherwise. The estimator is approximately unbiased if the cells are formed with equal response probability. However, the resulting estimator may not be efficient when \( y \)-values are not homogeneous within each cell. Eltinge & Yansaneh (1997), Da Silva & Opsomer (2004), and Smith et al. (2004) discussed this problem in detail.
The NWA methods discussed in this paper can provide a useful solution in this situation. If we use the regression NWA estimator with the auxiliary variable \( x_i = (x_{i1}, x_{i2}, \ldots, x_{iG-1}) \), the resulting regression estimator can be written

\[
\bar{y}_{re} = N^{-1} \sum_{g=1}^{G} \left( \sum_{i \in A} \pi_i^{-1} x_{ig} \right) \sum_{i \in A} \pi_i^{-1} \hat{p}_i^{-1} R_i x_{ig} y_i \sum_{i \in A} \pi_i^{-1} \hat{p}_i^{-1} R_i x_{ig}.
\] (34)

Here, we do not have to assume the equal response probability within each cell because the nonresponse bias is already taken care of by direct inclusion of the estimated response probability. Efficiency can be improved if the cells are formed with homogeneous \( y \)-values within each cell. Therefore, the regression NWA estimator in (34) controls the bias by using the estimated response probability and controls the variance by a suitable choice of cells. The price we pay for (34) is the additional computation in the variance estimation method incorporating the effect of using estimated response probability.

The NWA methods discussed so far assume that the response mechanism satisfies \([R.2]\) and \([R.3]\). As Kott (1994) and Särndal & Lundström (2006) have noted, assumption \([R.3]\) that \( p_i \) is bounded below can be unrealistic because most surveys have a fraction of hard core nonresponse. Assumption \([R.2]\) also implies that the response mechanism is ignorable in the sense of Rubin (1976). Alternative adjustment methods not requiring assumptions \([R.2]\)–\([R.3]\) can be a topic of future research.

**APPENDIX**

*Proof of Theorem 1.* To prove (12), we apply a Taylor expansion to get

\[
\bar{y}_e = \bar{y}_d + A_n (\hat{\alpha} - \alpha_0^{\alpha}) + 0.5(\hat{\alpha} - \alpha_0^{\alpha})^\top B_n (\hat{\alpha} - \alpha_0^{\alpha}),
\] (A.1)

where \( \hat{\alpha} \) is on the line segment joining \( \hat{\alpha} \) and \( \alpha_0^{\alpha} \). Thus, we can write

\[
\bar{y}_e = \bar{y}_d + A_n (\hat{\alpha} - \alpha_0^{\alpha}) + 0.5(\hat{\alpha} - \alpha_0^{\alpha})^\top B_n (\hat{\alpha} - \alpha_0^{\alpha}).
\]

where

\[
A_n = N^{-1} \sum_{i \in A} \pi_i^{-1} R_i \left( \frac{\partial p_i^{-1}}{\partial \alpha} \bigg|_{\alpha=\alpha_0^{\alpha}} \right) y_i,
\]

\[
B_n = \frac{1}{N} \sum_{i \in A} \pi_i^{-1} R_i \left( \frac{\partial^2 p_i^{-1}}{\partial \alpha \partial \alpha^\top} \bigg|_{\alpha=\hat{\alpha}} \right) y_i.
\]

From the definition of \( h_i \), \( \partial p_i^{-1}/\partial \alpha = p_i^{-1}(1 - p_i^{-1}) \), using \( \partial p_i^{-1}/\partial \alpha = -(p_i^{-1} - 1) \) and

\[
\frac{\partial^2 p_i^{-1}}{\partial \alpha \partial \alpha^\top} = p_i^{-1}(1 - p_i^{-1}) h_i h_i^\top - p_i^{-1}(1 - p_i^{-1}) \frac{\partial h_i}{\partial \alpha},
\]

we have

\[
A_n = N^{-1} \sum_{i \in A} \pi_i^{-1} R_i p_i^{-1}(p_i^{-1} - 1) h_i \partial \alpha y_i,
\]

and

\[
B_n = N^{-1} \sum_{i \in A} \pi_i^{-1} R_i p_i^{-1}(1 - p_i^{-1}) \left( h_i h_i^\top - \frac{\partial h_i}{\partial \alpha} \right) y_i.
\]
where $h_{i\lambda}$ is the value of $h_i(\alpha)$ evaluated at $\alpha = \bar{\alpha}$. Thus,

$$A_n = N^{-1} \sum_{i \in A} \pi_i^{-1}(p_{i|A} - 1)h_{i0}y_i + O_p(n^{-1/2})$$

and, under (4) and (10),

$$B_n = O_p(1). \quad \text{(A.2)}$$

Thus, using (11) and (A.2), we have

$$\bar{y}_e = \bar{y}_d + A_n^\top(\hat{\alpha} - \alpha_A^0) + O_p(n^{-1}). \quad \text{(A.3)}$$

Now, using

$$\frac{\partial}{\partial \alpha} \{ (R_i - p_{i|A})h_i \} = (R_i - p_{i|A}) \frac{\partial h_i}{\partial \alpha} + \left\{ \frac{\partial}{\partial \alpha} (R_i - p_{i|A}) \right\} h_i^\top$$

condition (11) can be written

$$\hat{\alpha} - \alpha_A^0 = \left\{ \sum_{i \in A} k_i p_{i|A}(1-p_{i|A})h_{i0}h_{i0}^\top \right\}^{-1} \sum_{i \in A} k_i (R_i - p_{i|A})h_{i0} + o_p(n^{-1/2}). \quad \text{(A.4)}$$

Inserting (A.4) into (A.3), we have (12).

ACKNOWLEDGEMENTS

This work was supported in part by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) and by a professional services contract from the U.S. National Center for Health Statistics. We thank Dong Wan Shin, Wayne A. Fuller, two anonymous referees, an Associate Editor, and the Editor for very constructive comments, which greatly improved this paper.

REFERENCES


Received 2 September 2005
Accepted 24 May 2007

Jae Kwang KIM: kim00020@yahoo.com
Department of Applied Statistics
Yonsei University, Seoul
Rep. of Korea, 120-749

Jay J. KIM: jkim5@cdc.gov
National Center for Health Statistics
Hyattsville, MD 20782, USA