Propensity score adjustment with several followups

Jae Kwang Kim* Jongho Im †

Abstract
Propensity score weighting adjustment is commonly used to handle unit nonresponse. When the response mechanism is nonignorable in the sense that the response probability depends directly on the study variable, a followup sample is commonly used to obtain an unbiased estimator using the framework of two-phase sampling, where the follow-up sample is assumed to respond completely. In practice, the followup sample is also subject to missingness. We consider propensity score weighting adjustment for nonignorable nonresponse when there are several follow-ups and the final follow-up sample is also subject to missingness. We propose two methods, one using calibration weighting and the other using a conditional likelihood using a so-called reverse conditional probability. Both methods provides consistent estimates under correct specification of the response model. A limited simulation study is used to compare the estimators. The proposed methods are applied to the real data example in a Korean household survey of employment.

Key Words: Two-phase sampling, Nonignorable nonresponse, Weighting, Survey Sampling

1. Introduction

Propensity score weighting method is a popular tool for handling unit nonresponse in survey sampling. Many surveys use the propensity score weighting method to reduce nonresponse bias (Fuller et al., 1994; Rizzo et al., 1996). If the responses are ignorable in the sense of Rubin (1976), then the propensity scores can be estimated consistently and the resulting propensity-score-adjusted (PSA) estimator is easily constructed. Kott (2006), Kim and Kim (2007), and Kim and Riddles (2012) has investigated some statistical properties of the PSA estimators under MAR case. If the responses are not ignorable, however, estimation of the propensity scores is complicated and often requires additional surrogate variables (Chen, Leung, and Qin, 2008) or instrumental variables (Kott and Chang, 2010) to estimate the model parameters consistently. Generally speaking, parameter estimation in the nonignorable response model can be subject to the identifiability problem and often requires additional assumptions (Wang et al, 2012).

Another way of handling nonignorable response model is to use followup samples to obtain further observations. Deming (1953) used two-phase sampling theory (Neyman 1938; Hansen and Hurwitz 1946) to obtain a followup sample in the nonrespondents' stratum and obtained a design-unbiased two-phase sampling estimator where the followup sample is treated as a second-phase sample in the two-phase sampling setup, assuming that the followup sample does not suffer from unit nonresponse. Proctor (1977) used a multinomial distribution to model differential response rate in the followup sample. Drew and Fuller (1980, 1981) extended the work of Proctor (1977) and developed a maximum likelihood estimation method for categorical response variable. Alho (1990) extended the approach of Drew and Fuller to the case of continuous response variable by adopting a logistic regression model for the response probability. Wood et al. (2006) used a fully parametric model to apply the EM algorithm.

In practice, we often have nonnegligible nonresponse even after several followup attempts. In the example of Korean Labor force survey discussed in Section 6, followup

*Department of Statistics, Iowa State University, Ames, IA 50011
†Department of Statistics, Iowa State University, Ames, IA 50011

attempts were made up to four times. After the fourth attempt, there are still about 10% nonrespondents in the sample. This paper, motivated by the Korean labor force survey example, proposed two estimation methods for handling nonresponse even after several followups. The proposed methods can be directly applicable to complex sampling setup. Up to the knowledge of the authors, the existing methods do not fully address nonresponse with several followup under complex sampling, with the exception of Drew and Fuller (1981) who only covered categorical survey items. Because we have several attempts, it turns out that we can not only estimate the model parameters consistently but also perform model diagnostics from the observed data.

Section 2 presents the basic setup. In Section 3, the first proposed method, based on the generalized method of moments technique for calibration weighting, is discussed. In Section 4, the second proposed method, based on conditional likelihood function, is presented. In Section 5, the proposed methods are compared in a limited simulation study. Real data application using Korean labor force survey is presented in Section 6. Concluding remarks are made in Section 7.

2. Basic setup

Let $U = \{1, 2, \cdots, N\}$ be the set of the finite population with known size N and $A(\subset U)$ be the original sample obtained from a probability sampling design. Let y_i be the study variable that can be obtained from the survey. Let d_i is the sampling weight assigned to unit i in the sample so that the resulting estimator

$$\hat{Y}_d = \sum_{i \in A} d_i y_i$$

is unbiased for the total $Y = \sum_{i=1}^{N} y_i$. We assume that the sampling weights satisfy $\sum_{i \in A} d_i = N$.

Now, suppose that the original sample is not fully observed and there are several followups to increase the number of respondents. Let $A_1(\subset A)$ be the set of initial respondents who provided answers to the surveys at the initial contact. Suppose that there are $T-1$ followups made to those who remain nonrespondents in the survey. Let $A_2(\subset A)$ be the set of respondents who provided answers to the surveys at the time of first followup. By definition, A_2 contains those already provided answers in the initial contact. Thus, $A_1 \subset A_2$. Similarly, we can define A_3 be the set of respondents who provided answers at the time of second followup. Continuing the process, we can define $A_1, \cdots, A_T$ such that

$$A_1 \subset \cdots \subset A_T.$$

Followup can be also called call-back. Suppose that there are T attempts (or $T-1$ followups) to obtain the survey response y_i and let δ_{it} be the response indicator function for y_i at the t-th attempt. If $\delta_{iT} = 0$, then the unit never responds and it is called hardcore nonresponse (Drew and Fuller, 1980). Using the definition of A_t, we can write $\delta_{it} = 1$ if $i \in A_t$ and $\delta_{it} = 0$ otherwise.

When the study variable y is categorical variable with K categories, Drew and Fuller (1980) proposed using a multinomial distribution with $T \times K + 1$ cells where the cell probabilities are defined by

$$\begin{aligned} \pi_{tk} &= \gamma(1-p_k)^{t-1} p_k f_k \\ \pi_0 &= (1-\gamma) + \gamma \sum_{k=1}^{K} (1-p_k)^T f_k \end{aligned}$$

where p_k is the response probability for category K , f_k is the population proportion such that $\sum_{k=1}^{K} f_k = 1$ and $1-\gamma$ is a proportion of hard-core nonrespondents. Thus, π_{tk} means the response probability that an individual in category k will respond at the t-th contact and π_0 is the probability that an individual will not have responded after T trials. Under simple random sampling, the maximum likelihood estimator of the parameter can be easily obtained by maximizing the log-likelihood

$$\log L = \sum_{t=1}^{T}\sum_{k=1}^{K} n_{tk}\log \pi_{tk} + n_0 \log \pi_0$$

where n_{tk} is the number of elements in the k-th category responding on the t-th contact and n_0 is the number of individual who did not respond up to T-th contact. Drew and Fuller (1981) further extended the results to complex survey sampling.

Alho (1990) considered the same problem with continuous y variable under simple random sampling. Alho (1990) defined p_{it} to be the conditional probability of $\delta_{it} = 1$, conditional on y_i and $\delta_{i,t-1} = 0$, and used the logistic regression model

$$p_{it} = P(\delta_{i,t} = 1|\delta_{i,t-1} = 0, x_i, y_i) = \frac{\exp(\alpha_t + x_i\phi_1 + y_i\phi_2)}{1+\exp(\alpha_t + x_i\phi_1 + y_i\phi_2)}, \quad t = 1, 2, \cdots, T, \tag{1}$$

for the conditional response probability where $\delta_{i0} \equiv 0$.

To estimate the parameters in (1), Alho (1990) also assumed that $(\delta_{i1}, \delta_{i2}-\delta_{i1}, \cdots, \delta_{iT}-\delta_{i,T-1}, 1-\delta_{iT})$ follows from a multinomial distribution with parameter $(\pi_{i1}, \pi_{i2}, \cdots, \pi_{iT}, 1-\sum_{t=1}^{T}\pi_{it})$ where $\pi_{it} = Pr\,(\delta_{i,t-1} = 0, \delta_{it} = 1 \mid x_i, y_i)$. Thus, we can write $\pi_{it} = p_{it}\prod_{k=1}^{t-1}(1-p_{ik})$. Under this setup, Alho (1990) considered maximizing the following conditional likelihood.

$$\begin{aligned} L(\phi) &= \prod_{\delta_{iT}=1}\prod_{t=2}^{T}\{Pr\,(\delta_{i1} = 1 \mid x_i, y_i, \delta_{iT} = 1)\}^{\delta_{i1}} \times \\ &\quad \{Pr\,(\delta_{it} = 1 \mid x_i, y_i, \delta_{i,t-1} = 0, \delta_{iT} = 1)\}^{\delta_{it}} \\ &= \prod_{\delta_{iT}=1}\left(\frac{\pi_{i1}}{1-\pi_{i,T+1}}\right)^{\delta_{i1}}\prod_{t=2}^{T}\left(\frac{\pi_{it}}{1-\pi_{i,T+1}}\right)^{\delta_{it}-\delta_{i,t-1}} \end{aligned} \tag{2}$$

where $\pi_{i,T+1} = 1 - \sum_{t=1}^{T}\pi_{it}$. To avoid the non-identifiability problem, Alho (1990) imposed

$$\sum_{i\in A/A_{t-1}} \delta_{it}\exp(-\alpha_t - \phi_1 x_i - \phi_2 y_i) = n - (n_1 + \cdots + n_t), \tag{3}$$

for $t = 1, 2, \cdots, T$. Note that (3) computes α_t given ϕ.

Alho's method used $\sum_{t=1}^{T}\hat{\pi}_{it} = 1 - \hat{\pi}_{i,T+1}$ to compute the propensity score-adjusted (PS) estimator

$$\hat{\theta}_{PS} = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_{iT}}{(1-\hat{\pi}_{i,T+1})}y_i. \tag{4}$$

Alho did not discuss variance estimation of the PS estimator in (4). Furthermore, Alho's method does not make use of auxiliary variable x_i in the nonrespondents and so there is still room for improvement.

3. Calibration weighting method

In this section, we propose an approach based on calibration weighting to estimate the model parameters in the conditional response model. Under the conditional response probability model in (1), we can compute $\tilde{\pi}_{it} = Pr(\delta_{it} = 1 \mid x_i, y_i)$ by $\tilde{\pi}_{it} = \sum_{j=1}^{t}\pi_{ij} =$

$\sum_{j=1}^{t}\{p_{ij}\prod_{k=1}^{j-1}(1-p_{ik})\}$. Thus, if y_i were observed, the model parameters $(\alpha_t, \phi_1, \phi_2)$ in the conditional response model (1) could be estimated by solving

$$\sum_{i\in A} d_i \delta_{it} \frac{1}{\tilde{\pi}_{it}}(1, x_i, y_i) = \sum_{i\in A} d_i (1, x_i, y_i), \tag{5}$$

for $(\alpha_t, \phi_1, \phi_2), t = 1, 2, \cdots, T$, where d_i is the sampling weight of unit i and. The equation (5) is often called calibration equation since the PSA estimator applied to $(1, x_i, y_i)$ leads to the full sample estimator. When the response mechanism is ignorable ($\phi_2 = 0$), calibration equation approach is quite intuitive and is quite popular in the propensity score weighting literature (Folsom, 1991; Iannacchione et al. 1991; Fuller et al., 1994; Kott, 2006; Kim and Riddles, 2012).

However, computing $\tilde{\pi}_{it}$ in (5) is somewhat complicated and the calibration equation is not easy to solve. Thus, instead of solving (5), one can use

$$\sum_{i\in A} d_i \delta_{i,t-1}(1, x_i, y_i) + \sum_{i\in A} d_i (1-\delta_{i,t-1}) \frac{\delta_{it}}{p_{it}}(1, x_i, y_i) = \sum_{i\in A} d_i (1, x_i, y_i) \tag{6}$$

for $t = 1, 2, \cdots, T$. Note that solving (6) is easier than solving (5) as (6) is a simple function of $(\alpha_t, \phi_1, \phi_2)$. In practice, neither (5) nor (6) can be used because the right side on the equality in (5) cannot be evaluated when y_i is missing for $\delta_{iT} = 0$.

To estimate the parameters, we can use the generalized method of moment (GMM) in a set of calibration equations that identifies the model parameters and also for the population parameters (X, Y). The calibration equations can be written as

$$\sum_{i\in A} d_i \delta_{i,t-1}(1, x_i, y_i) + \sum_{i\in A} d_i (1-\delta_{i,t-1}) \frac{\delta_{it}}{p_{it}}(1, x_i, y_i) = (N, X, Y), \tag{7}$$

for $t = 1, 2, \cdots, T$, and

$$\sum_{i\in A} d_i (1, x_i) = (N, X) \tag{8}$$

where $(\alpha_1, \cdots, \alpha_T, \phi_1, \phi_2)$ and (X, Y) are the unknown parameters to be determined. Thus, we have $p+q$ parameters ($p = \dim(x)$ and $q = \dim(y)$) with $(p+q)(T-1)$ equations. When $T > 1$, we have more equations than parameters and so we can apply the generalized method of moment (GMM) technique to compute the estimates. If we impose restrictions

$$\sum_{i\in A} d_i \delta_{i,t-1} + \sum_{i\in A} d_i (1-\delta_{i,t-1}) \frac{\delta_{it}}{p_{it}} = N, \quad t = 1, 2, \cdots, T, \tag{9}$$

then the resulting GMM estimation becomes a constrained GMM estimation.

Writing $\eta = (\alpha_1, \cdots, \alpha_T, \phi_1, \phi_2, X, Y)$, the GMM estimate can be obtained by minimizing

$$Q = \hat{U}^T(\eta) \left[\hat{V}\{\hat{U}(\eta)\}\right]^{-1} \hat{U}(\eta) \tag{10}$$

subject to (9), where $\hat{U}(\eta)$ is the system of equations defined in (7) and (8) and $\hat{V}\{\hat{U}(\eta)\}$ is a design-consistent variance estimator of $\hat{U}(\eta)$ for fixed value of η. Computational details of the constrained GMM are presented in Appendix A. Once the parameter ϕ is estimated from the GMM method, our final PS estimator is computed by

$$\hat{Y}_{PS} = \sum_{i\in A} d_i \frac{\delta_{iT}}{(1-\hat{\pi}_{i,T+1})} y_i. \tag{11}$$

For variance estimation, we propose using a replication method, such as the jackknife. Let $\hat{\eta}$ be the solution to the constrained GMM estimation. Under complete response, the replication variance estimator of $\hat{Y}_{HT} = \sum_{i \in A} d_i y_i$ can be constructed by

$$\hat{V}(\hat{Y}_{HT}) = \sum_{k=1}^{L} c_k \left(\hat{Y}_{HT}^{(k)} - \hat{Y}_{HT} \right)^2$$

where L is the size of the replication, c_k is the replication factor that is determined by the replication method, and $\hat{Y}_{HT}^{(k)} = \sum_{i \in A} d_i^{(k)} y_i$ is the k-replicate of $\hat{Y}_{HT}$ computed by $\hat{Y}_{HT}^{(k)} = \sum_{i \in A} d_i^{(k)} y_i$ for some replication weights $d_i^{(k)}$. See Chapter 4 of Fuller (2009) for more details.

Now, to construct the replication variance estimator of the PS estimator in (4), we first compute the replicate of $\hat{\eta}$, denoted by $\hat{\eta}^{(k)}$, from the same constrained GMM method using $d_i^{(k)}$ instead of d_i in $\hat{U}(\eta)$. Once $\hat{\eta}^{(k)}$ are obtained, the replication variance estimator of the PS estimator is computed by

$$\hat{V}(\hat{Y}_{PS}) = \sum_{k=1}^{L} c_k \left(\hat{Y}_{PS}^{(k)} - \hat{Y}_{PS} \right)^2$$

where

$$\hat{Y}_{PS}^{(k)} = \sum_{i \in A} d_i^{(k)} \frac{\delta_{iT}}{(1 - \hat{\pi}_{i,T+1}^{(k)})} y_i$$

and $\hat{\pi}_{i,T+1}^{(k)} = \hat{\pi}_{i,T+1}(\hat{\eta}^{(k)})$. Because $\hat{Y}_{PS}$ is a smooth function of $\hat{\eta}$, consistency of the replication variance estimator follows from the standard arguments.

4. Conditional maximum likelihood method

We now consider an alternative approach of computing the parameters in the propensity score model. The basic idea is to maximize the conditional likelihood among the set of respondents, those with $\delta_{i,T} = 1$, where the response probability is reversed in the sense that, instead of the original probability in (1), the conditional probability of $\delta_{i,t-1} = 1$ given that $\delta_{i,t} = 1$ is considered. The conditional likelihood was also considered by Tang et al (2003) and Pfeffermann and Sikov (2011) for the special case of $T = 1$, i.e. no followup.

The alternative approach based on conditional likelihood consists of two steps. In the first step, the reverse conditional probability $q_{it} = Pr\left(\delta_{it} = 1 \mid \delta_{i,t+1} = 1, x_i, y_i\right)$ is derived from the assumed response model. The reverse conditional probability is the conditional probability of response at time t given that it belongs to A_{t+1}. The reverse conditional probability can be derived from a Bayes formula. That is, we can obtain

$$q_{it} = O_{it}/(1 + O_{it}) \tag{12}$$

where

$$\begin{aligned} O_{it} &\equiv \frac{P(\delta_{it} = 1 \mid x_i, y_i, \delta_{i,t+1} = 1)}{P(\delta_{it} = 0 \mid x_i, y_i, \delta_{i,t+1} = 1)} \\ &= \frac{P(\delta_{it} = 1, \delta_{i,t+1} = 1 \mid x_i, y_i)}{P(\delta_{it} = 0, \delta_{i,t+1} = 1 \mid x_i, y_i)} \\ &= \frac{P(\delta_{i,t+1} = 1 \mid x_i, y_i, \delta_{i,t} = 1)}{P(\delta_{i,t+1} = 1 \mid x_i, y_i, \delta_{i,t} = 0)} \frac{P(\delta_{it} = 1 \mid x_i, y_i)}{P(\delta_{it} = 0 \mid x_i, y_i)} \\ &= \frac{1}{p_{i,t+1}} \frac{\tilde{\pi}_{it}}{1 - \tilde{\pi}_{it}} \end{aligned}$$

and $\tilde{\pi}_{it} = \sum_{j=1}^{t}\{p_{ij}\prod_{k=1}^{j-1}(1-p_{ik})\}$. Thus, we can express q_{it} as a function of $\alpha_t^* = (\alpha_1,\cdots,\alpha_t)$ and ϕ in (1).

In the second step, the parameter estimate is obtained by maximizing the conditional likelihood based on the reverse conditional probability. The conditional likelihood to be maximized at time t is

$$L_t(\alpha_t^*,\phi) = \prod_{i\in A_{t+1}} q_{it}^{\delta_{it}}\,(1-q_{it})^{1-\delta_{it}}\,,$$

where q_{it} is a function of $\alpha_t^* = (\alpha_1,\cdots,\alpha_t)$ and ϕ in (1). For the samples obtained from unequal probability sampling design, we can consider maximizing the pseudo conditional log-likelihood function given by

$$l(\alpha,\phi) = \sum_{t=1}^{T-1}\sum_{i\in A} d_i\delta_{i,t+1}\left\{\delta_{it}\log(q_{it}) + (1-\delta_{it})\log(1-q_{it})\right\}. \tag{13}$$

To incorporate the observed auxiliary information outside A_t, we add the following constraint

$$\sum_{i\in A} d_i(1-\delta_{i,t-1})\frac{\delta_{iT}}{\hat{p}_{it}} = \sum_{i\in A} d_i(1-\delta_{i,t-1}), \quad t=1,2,\cdots,T \tag{14}$$

$$\sum_{i\in A} d_i\frac{\delta_{iT}}{(1-\hat{\pi}_{i,T+1})}x_i = \sum_{i\in A} d_i x_i. \tag{15}$$

Incorporating the constraints into the PS estimation is equivalent to finding the solution that is the stationary point of the following Lagrangian function

$$L(\alpha,\phi,\lambda) = l(\alpha,\phi) + \lambda^T g(\alpha,\phi) \tag{16}$$

where $g(\alpha,\phi)$ are the constraint functions in (14) and (15). Once the parameters are estimated, we can use $\hat{Y}_{PS}$ in (11) to estimate the total Y. Computational details and the asymptotic normality of the resulting constrained pseudo maximum likelihood estimator are discussed in Appendix B.

Instead of deriving the reverse conditional probability q_{it} from the original response probability p_{it} in (1), one can directly assume a model for $q_{it} = Pr\,(\delta_{it}=1 \mid \delta_{i,t+1}=1, x_i, y_i)$. In this case, because y_i are observed when $\delta_{i,t+1}=1$, we can directly estimate the parameters from the maximum likelihood method. For example, we can directly construct a model

$$Pr\,(\delta_{it}=1 \mid \delta_{i,t+1}=1, x_i, y_i) = \frac{\exp(\alpha_t^* + \phi_1^* x_i + \phi_2^* y_i)}{1+\exp(\alpha_t^* + \phi_1^* x_i + \phi_2^* y_i)} := q_{it}(\alpha_t^*,\phi_1^*,\phi_2^*) \tag{17}$$

for $t = 1,2\cdots,T$, where it is understood that $\delta_{i,T+1}=1$ for all $i\in A$. In this case, the maximum likelihood for ϕ^* is obtained by maximizing the pseudo log-likelihood (13). To incorporate the observed auxiliary information of x_i outside A_T, we can impose additional constraint

$$\sum_{i\in A} d_i\frac{\delta_{i,T}}{q_{iT}}(1,x_i) = \sum_{i\in A} d_i(1,x_i)\,. \tag{18}$$

There are several advantages of using a direct model for q_{it} such as (17). First, the computation is easy and straightforward. Second, because we always observe (x_i,y_i) for $\delta_{i,T}=1$, model is easy to verify from the sample. That is, we can use the model diagnostic tools from the observed sample directly. Third, modeling itself is also flexible. Instead of assuming equal coefficient (ϕ_1^*, ϕ_2^*) in (17), we can build a more general model

$$Pr\left(\delta_{it}=1 \mid \delta_{i,t+1}=1, x_i, y_i\right) = \frac{\exp(\alpha_t^* + \phi_{1t}^* x_i + \phi_{2t}^* y_i)}{1+\exp(\alpha_t^* + \phi_{1t}^* x_i + \phi_{2t}^* y_i)} := q_{it}(\alpha_t^*, \phi_{1t}^*, \phi_{2t}^*) \tag{19}$$

for $t = 1, 2 \cdots, T-1$. Once a consistent estimator for (ϕ_{1t}, ϕ_{2t}) are constructed for $t = 1, \cdots, T-1$, we can perform a hypothesis testing for

$$H_0 : (\phi_{1t}, \phi_{2t}) = (\phi_1, \phi_2)\,.$$

If the null hypothesis cannot be rejected, then we can use model (17) for constructing the PS estimator. If the null hypothesis is rejected, then one may consider a model such as

$$\begin{pmatrix} \hat{\phi}_{1t} \\ \hat{\phi}_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} & \beta_{11} \\ \beta_{20} & \beta_{21} \end{pmatrix} \begin{pmatrix} 1 \\ t \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

and predict $(\hat{\phi}_{1T}, \hat{\phi}_{2T}) = (\hat{\beta}_{10} + \hat{\beta}_{11}T, \hat{\beta}_{20} + \hat{\beta}_{21}T)$ to obtain the PS estimator using $\hat{q}_{iT} = q_i(\hat{\phi}_{1T}, \hat{\phi}_{2T})$. See Section 6 for illustration of the pre-test estimation.

5. Simulation study

We perform a simulation study from an artificial finite population of size $N = 10,000$ with $x \sim N(0, 1/4)$ and $u \sim N(0, 1/4)$ where x, u are independent. In addition, we construct $y_i = x_i + u_i$ with $x_i - 1 \le y_i \le x_i + 1$. That is, y_i follows from a truncated normal distribution. From the finite population, we selected $B = 2,000$ independent Monte Carlo samples $\{(x_i, y_i), i = 1, 2 \cdots, n\}$ of size $n = 400$ using the simple random sampling.

From each sample, we applied two callbacks to obtain the followup samples. In each trial we have the same response probability, up to the intercept term α_t,

$$p_{it} = g(\alpha_t + 0.2x_i + 0.5y_i)$$

where α_t is determined to satisfy the desired response rate and $g(x) = \exp(x)/\{1 + \exp(x)\}$. Note that the response model is essentially the same model considered in Alho (1990). The overall response rate is about 38.2% for $t = 1$ and 72.6% for $t = 2$.

From the sample generated above, we computed four estimators of $\theta = E(Y)$: full sample estimator assuming no nonresponse, Alho's estimator (Alho), conditional maximum likelihood estimator (CMLE), and the calibration estimator (CAL). Table 1 presents the performance of the three estimators with followup samples. All the estimators have negligible biases but both CMLE and CAL are more efficient than Alho's estimator because they use extra auxiliary information (x_i) observed throughout the sample. The CMLE is slightly more efficient than the CAL estimator because the maximum likelihood estimator is generally more efficient than the method-of-moment estimator under the correct parametric model.

Variance estimators of calibration and conditional maximum likelihood methods were also computed. For calibration, we used the replication method discussed in Section 3 while we used the linearization method to compute the variance estimate for the conditional maximum likelihood method. Both methods show negligible biases in the simulation, less than 7 percent of the relative biases in absolute values.

Table 1: Monte Carlo biases, variances, and mean squared errors (MSE) of the point estimators.

Estimator	Bias	Variance	MSE
Full Sample	0.0000	0.0011	0.0011
Alho	0.0059	0.0032	0.0033
CMLE	0.0044	0.0021	0.0021
CAL	0.0031	0.0024	0.0024

Table 2: Response and nonresponse in 2009 Korean LALF survey

status	T=1	T=2	T=3	T=4	No reponse
Employment	81,685	46,926	28,124	15,992	
Unemployment	1,509	948	597	352	32350
Not in LF	57882	32308	19086	10790	

6. Application

The proposed methods were applied to the 2009 Korean Local-Area labor force (KLALF) survey. The KLALF is a large-scale labor force survey to get improved local-area level estimates. In the 2009 KLALF data, 157,205 sample households were contacted up to four followup. Table 2 displays the realized number of respondents for each of the follow-up attempts.

Table 3 shows the result of estimated ϕ_t for t=1,2,3 using logistic regression reverse propensity model

$$P(\delta_{it} = 1|\delta_{i,t+1} = 1, y_i) = \frac{\exp(\alpha_t + \phi_t y_i)}{1 + \exp(\alpha_t + \phi_t y_i)} \tag{20}$$

where $y_i = 1$ if unemployed and $y_i = 0$ otherwise. Because of the monotone (decreasing) structure of the missing data, y_i are observed when $\delta_{t+1} = 1$ and the parameters in (20) can be easily obtained by the maximum likelihood method. From the estimation results in Table 3, we concluded that ϕ_t is constant over follow-ups. Thus, we can safely assume that the reverse propensity model is

$$P(\delta_{it} = 1|\delta_{i,t+1} = 1, y_i) = \frac{\exp(\alpha_t + \phi y_i)}{1 + \exp(\alpha_t + \phi y_i)}. \tag{21}$$

We are interested in estimating three parameters θ_1, θ_2 and θ_3, which denote the proportion of employment, unemployment and not in labor force, respectively. Note that $\theta_3 = 1 - \theta_1 - \theta_2$ and so we report the result for θ_1 and θ_2 only. Under the assumed response model (21), we obtained five different estimates. The first one is the naive estimator that is computed by the simple mean of the respondents without making any adjustment. The other estimates are computed using Drew and Fuller (1980) method, Alho (1990) method and our two proposed methods. In computing Alho's method, we use the conditional probability model (1). We applied variance estimation method discussed in Section 3 and Section

Table 3: Estimated ϕ_t in model (20)

	$\hat{\phi}_t$	95% C.I. for ϕ_t
t=1	-0.112	(-0.191,-0.031)
t=2	-0.112	(-0.200,-0.025)
t=3	-0.110	(-0.219,-0.002)

Table 4: Estimated parameters for labor force in Korean LALF

Parameter	Method	Estimates	S.E($\times 10^{-3}$)
θ_1	Naive	0.5831	9.37
	Alho	0.5829	9.28
	Drew & Fuller	0.5847	10.06
	CMLE	0.5829	9.30
	Calibration	0.5830	9.33
θ_2	Naive	0.0115	2.11
	Alho	0.0119	2.70
	Drew & Fuller	0.0119	2.44
	CMLE	0.0119	2.31
	Calibration	0.0117	2.46

4 for the conditional maximum likelihood method and the calibration method. In computing Alho's method, we use the conditional probability model (1). For Drew and Fuller's model, we used an EM algorithm to parameter estimates for In Table 4, the four methods (excluding Naive method) produce slightly increased estimates for unemployment rate, which implies that the missing rate is higher for unemployed people, which is also verified from the result in Table 3. The CMLE shows the smallest estimated standard error among the methods considered.

7. Concluding remarks

We have considered the problem of parameter estimation under nonignorable nonresponse when the followup sample is also subject to missingness. Under the conditional response model (1), we can use a constrained maximum likelihood likelihood method to improve the efficiency by incorporating the auxiliary information, as discussed in Section 4, or use a calibration weighting method based on the constrained generalized method of moment method, as discussed in Section 3. Both methods effectively incorporate the auxiliary information available throughout the sample and provide consistent propensity-score-adjusted estimator under the conditional response model $p_{it} = Pr\left(\delta_{i,t} = 1 \mid \delta_{i,t-1} = 0, x_i, y_i\right)$. The price to pay is the computational complexity associated with the constrained optimization.

In large-scale survey data, computational complexity can be an issue and in this case a simple approximation can be used by employing a parametric model for the reverse conditional probability $q_{it} = Pr\left(\delta_{i,t} = 1 \mid \delta_{i,t+1} = 1, x_i, y_i\right)$. The approach based on the reverse conditional probability simplifies the computation and enables model diagnostics. For example, in the Korean Labor force data example in Section 6, the reverse propensity model is used to test whether the slope remains the same over the followup attempts. Such approach seems to be useful for the official surveys where the survey participation is no longer mandatory.

Appendix

A. Computation for Constrained GMM

We now discuss the computation for GMM method in Section 3. Let $X = \sum_{i=1}^{N} x_i$ and $Y = \sum_{i=1}^{N} y_i$. Writing $\hat{\theta}_t(x) = \sum_{i \in A} d_i \{\delta_{i,t-1} + (1 - \delta_{i,t-1})\delta_{it} p_{it}^{-1}\} x_i$, the calibration equation can be expressed as

$$\left(\hat{\theta}_t(1), \hat{\theta}_t(x), \hat{\theta}_t(y)\right) = (N, X, Y), \;\; t = 1, 2, \cdots, T$$

and

$$\left(\hat{\theta}_{HT}(1), \hat{\theta}_{HT}(x)\right) = (N, X).$$

Thus, writing $\eta = (N, X, Y, \alpha_1, \cdots, \alpha_T, \phi_1, \phi_2)$, we have

$$U(\eta) = \begin{bmatrix} U_1(\eta) \\ U_{HT}(\eta) \\ U_x(\eta) \\ U_y(\eta) \end{bmatrix}$$

where

$$\begin{aligned} U_1(\eta)' &= \left[\hat{\theta}_1(1) - N, \cdots, \hat{\theta}_T(1) - N\right] \\ U_{HT}(\eta)' &= \left[\hat{\theta}_{HT}(1) - N, \hat{\theta}_{HT}(x) - X\right] \\ U_x(\eta)' &= \left[\hat{\theta}_1(x) - X, \cdots, \hat{\theta}_T(x) - X\right] \\ U_y(\eta)' &= \left[\hat{\theta}_1(y) - Y, \cdots, \hat{\theta}_T(y) - Y\right]. \end{aligned}$$

The variance-covariance matrix of $U(\eta)$ are easily computed for fixed value of the parameter η by ignoring the randomness of δ's. The optimal value of η minimizing the Q-term can be obtained by minimizing

$$Q = U(\eta)' \{V(U)\}^{-1} U(\eta) \tag{A.1}$$

with respect to η.

However, the solution $\hat{\eta}$ may not satisfy $U_1(\hat{\eta}) = 0$ and $U_{HT}(\hat{\eta}) = 0$. We should impose the restriction $U_1(\hat{\eta}) = 0$ and $U_{HT}(\hat{\eta}) = 0$ into the GMM estimation. To discuss the constrained GMM, write (A.1) as

$$Q = \begin{pmatrix} U_a(\eta)' \\ U_b(\eta)' \end{pmatrix} \begin{Bmatrix} V_{aa} & V_{ab} \\ V_{ba} & V_{bb} \end{Bmatrix}^{-1} \begin{pmatrix} U_a(\eta) \\ U_b(\eta) \end{pmatrix} \tag{A.2}$$

where

$$U_a(\eta) = \begin{bmatrix} U_1(\eta) \\ U_{HT}(\eta) \end{bmatrix}$$

$$U_b(\eta) = \begin{bmatrix} U_x(\eta) \\ U_y(\eta) \end{bmatrix}$$

and obtain

$$Q = Q_a + Q_b$$

where

$$Q_a = U_a(\eta)' \{V_{aa}\}^{-1} U_a(\eta) \tag{A.3}$$

and

$$Q_b = \left(U_b - V_{ba}V_{aa}^{-1}U_a\right)' \left\{V_{bb} - V_{ba}V_{aa}^{-1}V_{ab}\right\}^{-1} \left(U_b - V_{ba}V_{aa}^{-1}U_1\right). \tag{A.4}$$

Thus, the constrained GMM with constraint $U_a(\hat{\eta}) = 0$ is equivalent to minimizing

$$Q_b^*(\eta_b) = U_b(\eta_b)' \left\{V_{bb} - V_{ba}V_{aa}^{-1}V_{ab}\right\}^{-1} U_b(\eta_b)$$

where η_b is the subvector of η excluding $\alpha_1, \cdots, \alpha_T$ and X.

We now discuss how to compute the variance-covariance matrix $V(U)$ in (A.1). Let an design-unbiased estimator of $\hat{\theta}_{HT}(y) = \sum_{i \in A} d_i y_i$ be of the form $\hat{V} = \sum_{i \in A} \sum_{j \in A} \Delta_{ij} y_i y_j$. To estimate the variance estimator of $\hat{\theta}_t(x) = \sum_{i \in A} d_i \eta_{it}$, where $\eta_{it} = \delta_{it} + (1 - \delta_{i,t-1})\delta_{it} x_i / p_{it}$, the naive variance estimator

$$\hat{V}_{naive,t} = \sum_{i \in A} \sum_{j \in A} \Delta_{ij} \eta_{it} \eta_{jt}$$

can be constructed, where $\Delta_{ij} = (\pi_{ij} - \pi_i \pi_j)/(\pi_{ij} \pi_i \pi_j)$. To see the unbiasedness of the naive variance estimator, note that

$$\begin{aligned}
E\left\{\hat{V}_{naive,t}\right\} &= E\left\{\sum_{i \in A} \sum_{j \in A} \Delta_{ij} E\left(\eta_{it}\eta_{jt} \mid A_{t-1}\right)\right\} \\
&= E\left\{\sum_{i \in A} \sum_{j \in A} \Delta_{ij} E\left(\eta_{it} \mid A_{t-1}\right) E\left(\eta_{jt} \mid A_{t-1}\right)\right\} \\
&\quad + E\left\{\sum_{i \in A} \sum_{j \in A} \Delta_{ij} Cov\left(\eta_{it}, \eta_{jt} \mid A_{t-1}\right)\right\} \\
&= E\left\{\sum_{i \in A} \sum_{j \in A} \Delta_{ij} x_i x_j\right\} + E\left\{\sum_{i \in A} \Delta_{ii} V\left(\eta_{it} \mid A_{t-1}\right)\right\} \\
&= V\left(\sum_{i \in A} d_i x_i\right) + E\left\{\sum_{i \in A} \left(\frac{1 - \pi_i}{\pi_i^2}\right) (1 - \delta_{i,t-1})(p_{it}^{-1} - 1) x_i^2\right\}.
\end{aligned}$$

On the other hand, the variance of $\hat{\theta}_t(x)$ is

$$\begin{aligned}
V\left(\hat{\theta}_t(x)\right) &= V\left\{E\left(\hat{\theta}_t(x) \mid A_{t-1}\right)\right\} + E\left\{V\left(\hat{\theta}_t(x) \mid A_{t-1}\right)\right\} \\
&= V\left(\sum_{i \in A} d_i x_i\right) + E\left\{\sum_{i \in A} \pi_i^{-2} \left(1 - \delta_{i,t-1}\right) (p_{it}^{-1} - 1) x_i^2\right\}.
\end{aligned}$$

Thus, ignoring the finite population correction term, the naive variance estimator is unbiased.

B. Constrained maximum likelihood method

Let $\theta \in \Theta = \{(\alpha, \phi); g(\alpha, \phi) = 0\}$. The solution $\hat{\theta}$ which maximizes (13) under constraints (14) and (15) is obtained by solving the following augmented Lagrangian function with additional parameters λ,

$$L(\theta, \lambda) = l(\theta) + \lambda^T g(\theta)$$

where $l(\theta)$ is defined in (13) and $g(\theta)$ corresponds the constraints in (14) and (15). Also the solution of this equation is equivalent to solve simultaneous non-linear equations $U_2(\theta, \lambda) = 0$, which is called Karush-Kuhn-Tucker(KKT) conditions (Boyd and Vandenberghe, 2004), where

$$U_2(\theta, \lambda) = \begin{bmatrix} g(\theta) \\ S(\theta) + \partial g(\theta)/\partial \theta^T \lambda \end{bmatrix}$$

where $S(\theta)$ is a score function for (13) such that

$$\begin{aligned} S(\theta) &= \sum_{t=2}^{T-1} \sum_{i \in A} h_{it} d_i \delta_{i,t+1} (\delta_{it} - q_{it}(\theta)) \\ h_{it} &= (1 + O_{it})^{-1} \partial \text{logit}(O_{it})/\partial \theta. \end{aligned}$$

If n_1 is the number of parameters for θ and n_2 is the number of constraints for $g(\theta)$, then we have $n_1 + n_2$ equations and $n_1 + n_2$ parameters for θ and λ in $U_2(\theta, \lambda)$. Thus, $U_2(\theta, \lambda) = 0$ has an exact solution and $\partial U_2(\theta, \lambda)/\partial(\theta, \lambda)$ is invertible under suitable assumptions.

Let $\eta = (\theta, \lambda)$ and define

$$\begin{aligned} U(\eta, Y) &\equiv \begin{bmatrix} U_1(\eta, Y) \\ U_2(\eta) \end{bmatrix} \\ U_1(\eta, Y) &= \hat{Y}_{cmle}(\eta) - Y \end{aligned}$$

where Y is population total and

$$\hat{Y}_{cmle}(\eta) = \sum_{i \in A} d_i \left\{ \delta_{iT} \left(1 - \pi_{i,T+1}(\theta)\right)^{-1} \right\} y_i.$$

Under some regularity conditions, by the Talyor linearization, the optimal estimator $\hat{Y}_{opt}$ can be written,

$$\hat{Y}_{opt} = \hat{Y}_{cmle}(\eta_0) + \frac{\partial U_1(\eta_0)}{\partial \eta}(\hat{\eta} - \eta_0) + o_p(n^{-1/2}) \quad \text{(B.1)}$$

$$U_2(\hat{\theta}^*) = U_2(\eta_0) + \frac{\partial U_2(\eta_0)}{\partial \eta}(\hat{\eta} - \eta_0) + o_p(n^{-1/2}) \quad \text{(B.2)}$$

Since the columns of $\partial U_2(\eta)/\partial \eta$ are linearly independent, $\partial U_2(\eta_0)/\partial \eta$ is invertible and, by substituting (B.2) to (B.1),

$$\hat{Y}_{opt} = \hat{Y}_{cmle}(\eta_0) - \frac{\partial U_1(\eta_0)}{\partial \eta} \left[\frac{\partial U_2(\eta_0)}{\partial \eta} \right]^{-1} U_2(\hat{\eta}) + o_p(n^{-1/2}). \quad \text{(B.3)}$$

and the asymptotic normality can follow by the standard argument.

REFERENCES

Alho, J.M. (1990), “Adjusting for nonresponse bias using logistic regression,” *Biometrika*, 77, 617–624.

Boyd, S. and Vandenberghe, L. (2004), *Convex Optimization*, Cambridge, UK:Cambridge University Press.

Chen, S. X., Leung, D. Y. H. and Qin, J. (2008), “Improved Semiparametric Estimation Using Surrogate Data,” *Journal of the Royal Statistical Society*, Ser. B, 70, 803–823.

Deming, W.E. (1953), “On a probability mechanism to attain an economic balance between the resultant error of response and the bias of nonresponse,” *Journal of the American Statistical Association*, 48, 743–772.

Drew, J.H. and Fuller, W.A. (1980),“ Modeling nonresponse in surveys with callbacks,” in *Proceedings of the Survey Research Section*, American Statistical Association, 639–642.

Drew, J.H. and Fuller, W.A. (1981),“ Nonresponse in complex multiphase surveys,” in *Proceedings of the Survey Research Section*, American Statistical Association, 623–628.

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

Fuller, W. A., Loughin, M. M., and Baker, H. D. (1994), “Regression weighting in the presence of nonresponse with application to the 1987-1988 nationwide food consumption survey,”*Survey Methodology*, 20, 75–85.

Fuller, W.A. (2009), *Sampling Statistics*, New Jersey:John Wiley & Sons, Inc.

Hansen, M.H., and Hurwitz, W.N. (1946), “The problem of nonresponse in sample surveys,”*Journal of the American Statistical Association*, 41, 517–529.

Iannacchione, V. G., Milne, J. G., and Folsom, R.E. (1991), “Response probability weight adjustments using logistics regression,” in *Proceedings of the Survey Research Methods Section*, American Statistical Association, 637–642.

Neyman, J. (1938), “Contribution to the theory of sampling human populations,” *Journal of the American Statistical Association*, 33, 101–116.

Kim, J.K. (2004), “Finite sample properties of multiple imputation estimators,” *The Annals of Statistics*, 32, 766–783.

Kim, J. K. and Kim, J. J. (2007), “Nonresponse weighting adjustment using estimated response probability,” *Canadian Journal of Statistics*, 35, 501–514.

Kim, J.K. and Riddles, M. (2012), “Some theory for propensity scoring adjustment estimator,” *Survey Methodology*, In press.

Kott,P.S. and Chang, T.(2010), “Using Calibration Weighting to Adjust for Nonignorable Unit Nonresponse,” *Journal of the American Statistical Association*, 105, 1265–1275.

Pfeffermann, D., and Sikov, A. (2011), “Imputation and estimation under nonignorable nonresponse in household surveys with missing covariate information,” *Journal of Official Statistics*, 27, 181–209.

Proctor, C. (1977), “Two direct approaches to survey nonreponse: estimating a proportion with callbacks and allocating effort to raise the response rate,” in *Proceedings of the Social Statistics Section*, American Statistical Association, 284–290.

Rizzo, L., Kalton, G., and Brick, J. M. (1996), “A comparison of some weighting adjustment methods for panel nonresponse,” *Survey Methodology*, 22, 44–53.

Tang, G., Little, R.J.A., and Raghunathan, T.E. (2003), “Analysis of Multivariate Missing Data with Nonignorable Nonresponse,” *Biometrika*, 90, 747–764.

Wang, S., Shao, J. and Kim, J.K.(2012), “An instrument variable approach for identification and estimation with Nonignorable Nonresponse,” *Submitted*.

Wood, A.M., White, I.R., and Hotopf, M. (2006),“Using number of failed contact attempts to adjust for nonignorable non-response,”*Journal of the Royal Statistical Society*, Ser. A, 169, 525–542.