Original Articles in statistics & probability

An Empirical Likelihood Approach to Reduce Selection Bias in Voluntary Samples

Calcutta Statistical Association Bulletin
75(1) 8–27, 2023

DOI: 10.1177/00080683231186488
journals.sagepub.com/home/csa
Sage

Jae Kwang Kim[1] and Kosuke Morikawa[2]

Abstract

How to construct the pseudo-weights in voluntary samples is an important practical problem in survey sampling. The problem is quite challenging when the sampling mechanism for the voluntary sample is allowed to be non-ignorable. Under the assumption that the sample participation model is correctly specified, we can compute a consistent estimator of the model parameter and construct the propensity score estimator of the population mean. We propose using the empirical likelihood method to construct the final weights for voluntary samples by incorporating the bias calibration constraints and the benchmarking constraints. Linearization variance estimation of the proposed method is developed. A toy example is also presented to illustrate the idea and the computational details. A limited simulation study is also performed to evaluate the performance of the proposed methods.

Keywords

non-ignorable non-response, missing not at random, propensity score estimation

AMS subject classification: 62D10, 63D05

1. Introduction

Probability sampling is a gold-standard method for obtaining a representative sample from a target population. Probability sampling allows constructing valid statistical inferences for finite population parameters. Classical approaches in survey sampling are discussed in Cochran[5], Särndal et al.[25], Fuller[12], and Tillé[27].

Despite the promise of probability samples, non-probability samples are common even though an appropriate representation of the target population is not guaranteed. Nowadays, collecting a strict probability sample is almost impossible due to unavoidable issues such as frame

[1]Department of Statistics, Iowa State University, Ames, Iowa, USA
[2]Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka, Japan

Corresponding author:
Jae Kwang Kim, Department of Statistics, Iowa State University, Ames, Iowa 50011, USA.
E-mail: jkim@iastate.edu

undercoverage and non-response. Also, the cost of strict probability sampling is increasing. However, statistical analysis of non-probability survey samples faces many challenges, as documented by Baker et al.[1]. Non-probability samples are subject to selection biases, and they do not represent the target population. A popular framework in dealing with biased non-probability samples is calibration weighting incorporating the auxiliary information observed throughout the finite population. Such calibration weighting method is based on the assumption that the sampling mechanism for the non-probability sample is ignorable after adjusting for the auxiliary variables used for calibration weighting. Such an assumption is essentially the missing at random (MAR) assumption of Rubin[24]. Using MAR assumptions, calibration weighting methods for non-probability samples have been discussed in Dever and Valliant[8] and Elliott and Valliant[9], among others.

However, the MAR assumption may not be satisfied in many cases. That is, the selection bias may exist even after controlling on the auxiliary variables that are observed throughout the finite population. In that case, we need to build an explicit model for the non-ignorable sampling mechanism that reflects the dependency of the sampling mechanism on the study variable that is subject to missingness. Under a correctly specified model of the sampling mechanism, the generalized method of moments approach in Kott and Chang[17] and Wang et al.[29] or the maximum likelihood estimation approach considered in Pfeffermann and Sikov[20], Riddles et al.[23], or Morikawa et al.[19] can be developed. Chapter 8 of Kim and Shao[14] contains an extensive review of statistical methods for non-ignorable non-response models.

In this article, motivated by Qin et al.[22], we consider combining non-ignorable selection model with empirical likelihood method to develop a unified approach to propensity score weighting for voluntary samples. Under model identification condtions, the proposed method can handle non-ignorable selection model. The final weights can incorporate the auxiliary variables through calibration weighting. Statistical inference with the final propensity score weighting is somewhat complicated due to the several steps in the final weighting. While the idea of using a non-ignorable response model to construct the final propensity weights for handling voluntary sample is natural, the literature on this research direction is somewhat limited. Thus, we present a systematic approach to using a non-ignorable sample participation model to adjust the selection bias and make inferences from a non-probability sample. The empirical likelihood is nicely applied to create the final weights. The proposed method, however, is based on the model assumption for the sampling mechanism. We also present a toy example to illustrate the idea and give the computational details. If a single model assumption is not feasible, we can compute several estimates under different model scenarios and may consider sensitivity analysis (Copas and Eguchi[6]).

The article is organized as follows. In Section 2, the basic setup and the research problem are introduced. In Section 3, propensity score estimation method under non-ignorable model is discussed. In Section 4, the final propensity score weighting method using empirical likelihood is proposed and its variance estimation is discussed. In Section 5, an illustrative example is used to present the computational detains of the proposed method. In Section 6, results from a limited simulation study are presented. An extension to a semiparametric non-ignorable selection model is discussed in Section 7. Some concluding remarks are made in Section 8.

2. Basic Setup

Let $U = \{1, \cdots, N\}$ be the index set of the finite population of size N. Let $S \subset U$ be the index set of sample. Let δ_i be the sample selection indicator of unit i such that $\delta_i = 1$ if $i \in S$ and $\delta_i = 0$

otherwise. We observe y_i only when $\delta_i = 1$. We assume that the vectors of auxiliary variables $\mathbf{x}_i$ are available throughout the finite population. We are interested in estimating the population mean $\theta = N^{-1}\sum_{i=1}^{N} y_i$ from the sample.

Let $\pi(\mathbf{x}, y) = P(\delta = 1 \mid \mathbf{x}, y)$ be the propensity score (PS) function for the sampling mechanism. If $\pi(\mathbf{x}, y)$ is known, the empirical likelihood (EL) method can be used to estimate the parameter of interest. That is, we first find the maximizer of

$$\ell(p) = \sum_{i\in S} \log(p_i)$$

subject to $\sum_{i\in S} p_i = 1$ and

$$\sum_{i\in S} p_i \pi(\mathbf{x}_i, y_i) = W, \tag{2.1}$$

where $W = P(\delta = 1)$ is the marginal probability of $\delta = 1$. If $W = N^{-1}\sum_{i=1}^{N} \pi_i$ is unknown, we can estimate W by the profile empirical likelihood method, as considered by Qin et al.[22], or simply use $\widehat{W}_{HT} = n/N$, where $n = |S|$ is the realized sample size. Constraint (2.1) can be called the (internal) bias calibration constraint (Firth and Bennett[10]). In addition, we may impose

$$\sum_{i\in S} p_i \mathbf{x}_i = \bar{\mathbf{X}}_N \tag{2.2}$$

as an additional restriction, where $\bar{\mathbf{X}}_N = N^{-1}\sum_{i=1}^{N} \mathbf{x}_i$. Constraint (2.2) can be called the benchmarking constraint. Constraint (2.1) is used to remove the selection bias and constraint (2.2) is used to improve the efficiency of the resulting EL estimator. The implicit assumption for using (2.2) is that the outcome model is linear as in $E(Y \mid \mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$ for some $\boldsymbol{\beta}$. If the outcome regression model is non-linear, we can directly use the model calibration of Wu and Sitter[30] to replace the benchmarking constraint in (2.2).

Once $\hat{p}_i$ are obtained from the above optimization, the final estimator of θ is obtained by

$$\hat{\theta}_{\text{EL}} = \sum_{i\in S} \hat{p}_i y_i. \tag{2.3}$$

The final estimator in (2.3) is often called the maximum empirical likelihood estimator (MELE) of θ. The above empirical likelihood estimator has been considered by Qin et al.[22], Kim[13], and Berger and Torres[2], among others.

If $W = N^{-1}\sum_{i=1}^{N} \pi_i$ is known, using the standard linearization, we can obtain

$$\hat{\theta}_{\text{EL}} = \frac{1}{N}\sum_{i=1}^{N} \left\{ \hat{y}_i + \frac{\delta_i}{\pi_i}(y_i - \hat{y}_i) \right\} + o_p(n^{-1/2}), \tag{2.4}$$

where $\hat{y}_i = \pi_i \hat{\beta}_1 + \mathbf{x}_i' \hat{\boldsymbol{\beta}}_2$ and

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = \left\{ \sum_{i \in S} \pi_i^{-2} \begin{pmatrix} \pi_i - W \\ \mathbf{x}_i - \bar{\mathbf{X}}_N \end{pmatrix} \begin{pmatrix} \pi_i - W \\ \mathbf{x}_i - \bar{\mathbf{X}}_N \end{pmatrix}' \right\}^{-1} \sum_{i \in S} \pi_i^{-2} \begin{pmatrix} \pi_i - W \\ \mathbf{x}_i - \bar{\mathbf{X}}_N \end{pmatrix} y_i .$$

A sketched proof for (2.4) is presented in Appendix A. Result (2.4) means that the EL estimator is asymptotically equivalent to a version of regression estimator which satisfies the calibration constraints.

The linearization formula in (2.4) is particularly useful in developing linearization variance estimation. That is, we can use

$$\hat{V} = \frac{1}{N(N-1)} \sum_{i=1}^{N} \left(\hat{\eta}_i - \bar{\eta}_N\right)^2 , \tag{2.5}$$

where $\hat{\eta}_i = \hat{y}_i + \delta_i \pi_i^{-1} (y_i - \hat{y}_i)$ and $\bar{\eta}_N = N^{-1} \sum_{i=1}^{N} \hat{\eta}_i$, to estimate the variance of $\hat{\theta}_{\mathrm{EL}}$. The linearization variance estimator in (2.5) is derived based on the assumption that δ_i are mutually independent to each other. If they are correlated as in survey sampling, the joint probabilities of (δ_i, δ_j) are needed to compute more accurate variance estimation.

3. Propensity Score Estimation Under a Parametric PS Model

In probability sampling, π_i are known and the EL method in Section 2 can be directly applicable. In the voluntary sampling, we do not know the propensity score function $\pi(\mathbf{x}, y)$. Instead, we may use a model on $\pi(\mathbf{x}, y)$, say $\pi(\mathbf{x}, y) = \pi(\mathbf{x}, y; \phi_0)$ for some ϕ_0, and develop methods for bias adjustment under the model. We first discuss how to estimate ϕ_0 from the voluntary sample and then discuss estimation of $\theta = E(Y)$.

We assume that the propensity score function follows a parametric model such that $\pi(\mathbf{x}, y) = \pi(\mathbf{x}, y; \phi_0)$ for some $\phi_0 \in \Phi \subset \mathbb{R}^q$. The observed likelihood function of ϕ derived from the marginal density of $(\delta_i, \delta_i y_i)$ given $\mathbf{x}_i$ can be written as

$$L_{\mathrm{obs}}(\phi) = \prod_{i=1}^{N} \left\{ f(y_i \mid \mathbf{x}_i) \pi(\mathbf{x}_i, y_i; \phi) \right\}^{\delta_i} \left\{ 1 - \tilde{\pi}(\mathbf{x}_i; \phi) \right\}^{1-\delta_i} \tag{3.1}$$

where

$$\tilde{\pi}(\mathbf{x}; \phi) = \mathbb{E} \left\{ \pi(\mathbf{x}, Y; \phi) \mid \mathbf{x} \right\} .$$

Thus, to construct the observed likelihood function in (3.1), we may need to make a model assumption about $f(y \mid \mathbf{x})$. Unfortunately, the resulting maximum likelihood estimator (MLE) of ϕ_0 is not robust against model misspecification of the outcome model (Copas and Li[7]). Because we do not observe y_i among $\delta_i = 0$, we cannot directly apply the model diagnostic tools for complete response.

Instead of using a model assumption for $f(y \mid \mathbf{x})$, we can make a model assumption for $f(y \mid \mathbf{x}, \delta = 1)$ with confidence, as $(\mathbf{x}_i, y_i)$ are observed for $\delta_i = 1$. Thus, we can safely assume that $f_1(y \mid \mathbf{x}) \equiv f(y \mid \mathbf{x}, \delta = 1)$ is correctly specified. In this case, we can use the following identity which was originally proved by Pfeffermann and Sverchkov[21]:

$$\tilde{\pi}(\mathbf{x}; \phi) = \left[\int \{\pi(\mathbf{x}, y; \phi)\}^{-1} f_1(y \mid \mathbf{x}) dy \right]^{-1}. \tag{3.2}$$

Using (3.2), we can construct

$$\ell_{\text{obs}}(\phi) = \sum_{i=1}^{N} \left\{ \delta_i \log \pi(\mathbf{x}_i, y_i; \phi) + (1 - \delta_i) \log \left(1 - \hat{\pi}(\mathbf{x}_i; \phi)\right) \right\}, \tag{3.3}$$

where

$$\hat{\pi}(\mathbf{x}; \boldsymbol{\phi}) = \left[\int \{\pi(\mathbf{x}, y; \boldsymbol{\phi})\}^{-1} \hat{f}_1(y \mid \mathbf{x}) dy \right]^{-1} \tag{3.4}$$

and $\hat{f}_1(y \mid \mathbf{x})$ is a consistent estimator of $f_1(y \mid \mathbf{x})$. If we define $\omega(\mathbf{x}, y; \boldsymbol{\phi}) = \{\pi(\mathbf{x}, y; \boldsymbol{\phi})\}^{-1}$ and $\widehat{\omega}(\mathbf{x}; \boldsymbol{\phi}) = \{\hat{\pi}(\mathbf{x}; \boldsymbol{\phi})\}^{-1}$, then (3.4) can be expressed as

$$\widehat{\omega}(\mathbf{x}; \boldsymbol{\phi}) = \int \omega(\mathbf{x}, y; \phi) \hat{f}_1(y \mid \mathbf{x}) dy. \tag{3.5}$$

The propensity weight in (3.5) can be called the smoothed propensity weight (Kim and Skinner[15]). See Section 5 for some computational details of the smoothed propensity score function. Furthermore, we impose

$$\sum_{i=1}^{N} \frac{\delta_i}{\hat{\pi}(\mathbf{x}_i; \phi)} = N \tag{3.6}$$

as a constraint for estimating the model parameter in the PS function. Including constraint (3.6) into the ML estimation will improve the efficiency of the final estimation. It was originally proposed by Cao et al.[3].

Morikawa et al.[19] show that the MLE can be obtained by solving

$$\sum_{i=1}^{N} \left\{ \frac{\delta_i}{\pi(\mathbf{x}_i, y_i; \phi)} - 1 \right\} \mathbf{b}^*(\mathbf{x}_i; \phi) = 0, \tag{3.7}$$

where

$$\mathbf{b}^*(\mathbf{x}; \phi) = E\left\{ \mathbf{h}(\mathbf{x}_i, Y; \phi) \pi(\mathbf{x}_i, Y; \phi) \mid \mathbf{x}_i, \delta_i = 0 \right\},$$
$$\mathbf{h}(\mathbf{x}, y; \phi) = \frac{\partial}{\partial \phi} \text{logit}\{\pi(\mathbf{x}, y; \phi)\},$$

and

$$O(\mathbf{x}, y; \phi) = \frac{1 - \pi(\mathbf{x}, y; \phi)}{\pi(\mathbf{x}, y; \phi)}.$$

More generally, we can estimate ϕ by solving

$$\hat{U}_b(\phi) \equiv \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{\delta_i}{\pi(\mathbf{x}_i, y_i; \phi)} - 1 \right\} \mathbf{b}(\mathbf{x}_i; \phi) = 0 \tag{3.8}$$

for some $\mathbf{b}(\mathbf{x}; \phi)$ such that the solution exists uniquely. For example, we can use $\mathbf{b}(\mathbf{x}_i; \phi) = \mathbf{x}_i$. Parameter estimation using (3.8) under a parametric PS model was first considered by Kott and Chang[17] and further explored by Wang et al.[29]. Note that we can achieve (3.6) by including the intercept term in $\mathbf{b}(\mathbf{x}; \phi)$. If $\hat{\phi}$ is obtained from (3.8), we can compute $\hat{\pi}_i = \pi(\mathbf{x}_i, y_i; \hat{\phi})$ in the sample and we can estimate $\theta = E(Y)$ by

$$\hat{\theta}_{\text{PS}} = \frac{1}{N} \sum_{i \in S} \frac{1}{\hat{\pi}_i} y_i. \tag{3.9}$$

Using the standard linearization, we can show that

$$\hat{\theta}_{\text{PS}} = \frac{1}{N} \sum_{i=1}^{N} \left\{ \boldsymbol{\gamma}' \mathbf{b}_i + \frac{\delta_i}{\pi_i} \left(y_i - \boldsymbol{\gamma}' \mathbf{b}_i \right) \right\} + o_p(n^{-1/2}), \tag{3.10}$$

where $\mathbf{b}_i = \mathbf{b}(\mathbf{x}_i; \phi)$ is defined in (3.8),

$$\boldsymbol{\gamma}' = E\left\{ Y \mathbf{h}_0(X, Y)' \right\} \left[E\left\{ \mathbf{b}(X) \mathbf{h}_0(X, Y)' \right\} \right]^{-1},$$

and $\mathbf{h}_0(X, Y) = \mathbf{h}(X, Y)\{1 - \pi(X, Y)\}$. A sketched proof for (3.10) is presented in Appendix B.

The linearization formula in (3.10) can be used to construct the linearized variance estimation. We can apply the same linearization formula in (2.5) with a different formula for $\hat{\eta}_i$. Since we use a different linearization result in (3.10), we may use

$$\hat{\eta}_i = \hat{\mathbf{b}}_i' \hat{\boldsymbol{\gamma}} + \frac{\delta_i}{\hat{\pi}_i} \left(y_i - \hat{\mathbf{b}}_i' \hat{\boldsymbol{\gamma}} \right)$$

in applying the variance estimation formula (2.5), where

$$\hat{\boldsymbol{\gamma}}' = \left(\sum_{i \in S} (\hat{\pi}_i^{-1} - 1) \hat{\mathbf{h}}_i' y_i \right) \left\{ \sum_{i \in S} (\hat{\pi}_i^{-1} - 1) \hat{\mathbf{b}}_i \hat{\mathbf{h}}_i' \right\}^{-1}$$

and ϕ is replaced by $\hat{\phi}$ in the linearization formula.

Now, to incorporate the auxiliary information, we consider the class of the regression PS estimator

$$\hat{\theta}_{RPS}(\boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N}\left\{\mathbf{x}_i'\boldsymbol{\beta} + \frac{\delta_i}{\hat{\pi}_i}\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)\right\} \tag{3.11}$$

where $\boldsymbol{\beta}$ is to be determined.

If we choose $\hat{\boldsymbol{\beta}}^*$ such that the asymptotic variance of $\hat{\theta}_{RPS}(\boldsymbol{\beta})$ is minimized among the class in (3.11), we obtain the optimal PS estimator. The optimal $\hat{\boldsymbol{\beta}}^*$ can be written as

$$\hat{\boldsymbol{\beta}}^* = \left[\widehat{V}\left\{\sum_{i\in S}\hat{\pi}_i^{-1}\mathbf{x}_i\right\}\right]^{-1}\widehat{Cov}\left\{\sum_{i\in S}\hat{\pi}_i^{-1}\mathbf{x}_i, \sum_{i\in S}\hat{\pi}_i^{-1}y_i\right\}$$

and we can use Taylor expansion to derive the explicit formula for the variance–covariance matrix. When $\hat{\phi}$ is obtained from (3.8), we can compute the optimal estimator by solving

$$\sum_{i\in S}\hat{\pi}_i^{-1}(1-\hat{\pi}_i)\left(y_i - \mathbf{x}_i'\boldsymbol{\beta} - \mathbf{b}_i'\boldsymbol{\gamma}\right)\mathbf{x}_i = 0$$

and

$$\sum_{i\in S}\hat{\pi}_i^{-1}(1-\hat{\pi}_i)\left(y_i - \mathbf{x}_i'\boldsymbol{\beta} - \mathbf{b}_i'\boldsymbol{\gamma}\right)\hat{\mathbf{h}}_i = 0$$

for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Morikawa and Kim[18] proposed an adaptive optimal estimator achieving the semiparametric lower bound.

4. Empirical Likelihood Method

Using the parameter estimation in Section 3, we can apply the EL method for final PS weighting. That is, we first find the maximizer of

$$\ell(p) = \sum_{i\in S}\log(p_i)$$

subject to $\sum_{i\in S} p_i = 1$,

$$\sum_{i\in S} p_i\pi(\mathbf{x}_i, y_i; \hat{\phi}) = N^{-1}\sum_{i=1}^{N}\hat{\pi}(\mathbf{x}_i; \hat{\phi}), \tag{4.1}$$

and

$$\sum_{i \in S} p_i \mathbf{x}_i = N^{-1} \sum_{i=1}^{N} \mathbf{x}_i, \tag{4.2}$$

where $\hat{\pi}(\mathbf{x}_i; \hat{\phi})$ is defined in (3.4) and $\hat{\phi}$ is computed from (3.8). Condition (4.1) is the bias calibration condition.

Note that the final weights are obtained in two steps. In the first step, a consistent estimator $\hat{\phi}$ of the model parameter in the PS model $\pi(\mathbf{x}, y; \phi)$ is computed. In the second step, we treat $\hat{\pi}_i = \pi(\mathbf{x}_i, y_i; \hat{\phi})$ as if the true inclusion probability and apply the EL method incorporating the bias-calibration constraints and the calibration constraint.

Now, to discuss the asymptotic property of the final EL estimator $\hat{\theta}_{\text{EL}}$, we apply the same two-step procedure to obtain Taylor linearization. In the first step, ignoring the uncertainty in $\hat{\phi}$ for now, we can apply the linearization method for obtaining (2.4) to get

$$\hat{\theta}_{\text{EL}} = \frac{1}{N} \sum_{i=1}^{N} \left\{ \hat{y}_i^{(0)} + \frac{\delta_i}{\pi(\mathbf{x}_i, y_i; \hat{\phi})} \left(y_i - \hat{y}_i^{(1)} \right) \right\} + o_p(n^{-1/2}), \tag{4.3}$$

where $\hat{y}_i^{(0)} = \hat{\pi}(\mathbf{x}_i; \hat{\phi})\hat{\beta}_1 + \tilde{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}_2$, $\hat{y}_i^{(1)} = \pi(\mathbf{x}_i, y_i; \hat{\phi})\hat{\beta}_1 + \mathbf{x}_i' \hat{\boldsymbol{\beta}}_2$ and

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = \left\{ \sum_{i \in S} \hat{\pi}_i^{-2} \begin{pmatrix} \hat{\pi}_i - \hat{W} \\ \mathbf{x}_i - \bar{\mathbf{X}}_N \end{pmatrix} \begin{pmatrix} \hat{\pi}_i - \hat{W} \\ \mathbf{x}_i - \bar{\mathbf{X}}_N \end{pmatrix}' \right\}^{-1} \sum_{i \in S} \hat{\pi}_i^{-2} \begin{pmatrix} \hat{\pi}_i - \hat{W} \\ \mathbf{x}_i - \bar{\mathbf{X}}_N \end{pmatrix} y_i.$$

Here, $\hat{\pi}_i = \pi(\mathbf{x}_i, y_i; \hat{\phi})$ and $\hat{W} = N^{-1} \sum_{i=1}^{N} \pi(\mathbf{x}_i; \hat{\phi})$.

Note that we can express (4.3) as

$$\hat{\theta}_{\text{EL}} = \frac{1}{N} \sum_{i=1}^{N} \left\{ \hat{\beta}_1 \left(\hat{\pi}(\mathbf{x}_i; \hat{\phi}) - \delta_i \right) + \mathbf{x}_i' \hat{\boldsymbol{\beta}}_2 + \frac{\delta_i}{\hat{\pi}_i} \left(y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_2 \right) \right\} + o_p(n^{-1/2}). \tag{4.4}$$

In the second step, we need to take into account the uncertainty in $\hat{\phi}$. To do this, we apply the Taylor expansion with respect to ϕ and obtain the final influence function.

Now, to apply the Taylor expansion with respect to ϕ, define

$$\begin{aligned} \hat{\theta}_\ell(\hat{\phi}) &= \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_1 \left(\hat{\pi}(\mathbf{x}_i; \hat{\phi}) - \delta_i \right) + \frac{1}{N} \sum_{i=1}^{N} \left\{ \mathbf{x}_i' \hat{\boldsymbol{\beta}}_2 + \frac{\delta_i}{\hat{\pi}_i} \left(y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_2 \right) \right\} \\ &:= \hat{\theta}_{\ell,1}(\hat{\phi}) + \hat{\theta}_{\ell,2}(\hat{\phi}) \end{aligned}$$

and note that (4.3) can be written as

$$\hat{\theta}_{\text{EL}} = \hat{\theta}_\ell(\hat{\phi}) + o_p(n^{-1/2}).$$

We can apply Taylor linearization to $\hat{\theta}_{\ell,1}(\hat{\phi})$ with respect to ϕ to get

$$\begin{aligned}\hat{\theta}_{\ell,1}(\hat{\phi}) &= \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_1\left(\hat{\pi}(\mathbf{x}_i;\phi)-\delta_i\right)+E\left\{N^{-1}\sum_{i=1}^{N}\hat{\beta}_1\frac{\partial}{\partial\phi'}\hat{\pi}(\mathbf{x}_i;\phi)\right\}\left(\hat{\phi}-\phi\right)+o_p(n^{-1/2})\\ &= \frac{1}{N}\sum_{i=1}^{N}\left[\hat{\beta}_1\{\hat{\pi}(\mathbf{x}_i;\phi)+\boldsymbol{\kappa}_1'\mathbf{b}_i\}-\frac{\delta_i}{\pi_i}\hat{\beta}_1\left\{\pi_i+\boldsymbol{\kappa}_1'\mathbf{b}_i\right\}\right]+o_p(n^{-1/2}),\end{aligned} \tag{4.5}$$

where $\pi_i=\pi(\mathbf{x}_i, y_i;\phi)$, $\boldsymbol{\kappa}_1'=-E\left\{\hat{\mathbf{g}}_1(X;\phi)'\right\}\left[E\left\{\mathbf{b}(X)\mathbf{h}_0(X, Y)'\right\}\right]^{-1}$ and

$$\begin{aligned}\hat{\mathbf{g}}_1(\mathbf{x};\phi) &= \frac{\partial}{\partial\boldsymbol{\phi}}\hat{\pi}(\mathbf{x};\boldsymbol{\phi})\\ &= \{\hat{\pi}(\mathbf{x};\boldsymbol{\phi})\}^2\left[\int\left\{\frac{1}{\pi(\mathbf{x}, y;\boldsymbol{\phi})}-1\right\}\mathbf{h}(\mathbf{x}, y;\phi)\hat{f}_1(y\mid\mathbf{x})dy\right].\end{aligned} \tag{4.6}$$

Also, we can apply the Taylor linearization to $\hat{\theta}_{\ell,2}(\hat{\phi})$ with respect to ϕ to get

$$\begin{aligned}\hat{\theta}_{\ell,2}(\hat{\phi}) &= \frac{1}{N}\sum_{i=1}^{N}\left\{\mathbf{x}_i'\hat{\boldsymbol{\beta}}_2+\frac{\delta_i}{\pi_i}\left(y_i-\mathbf{x}_i'\hat{\boldsymbol{\beta}}_2\right)\right\}\\ &\quad -E\left\{N^{-1}\sum_{i=1}^{N}\delta_i\pi_i^{-2}\pi_i(1-\pi_i)(y_i-\mathbf{x}_i'\hat{\boldsymbol{\beta}}_2)\mathbf{h}_i'\right\}\left(\hat{\phi}-\phi\right)+o_p(n^{-1/2})\\ &= \frac{1}{N}\sum_{i=1}^{N}\left\{\mathbf{x}_i'\hat{\boldsymbol{\beta}}_2+\mathbf{b}_i'\boldsymbol{\kappa}_2+\frac{\delta_i}{\pi_i}\left(y_i-\mathbf{x}_i'\hat{\boldsymbol{\beta}}_2-\mathbf{b}_i'\boldsymbol{\kappa}_2\right)\right\}+o_p(n^{-1/2}),\end{aligned} \tag{4.7}$$

where

$$\boldsymbol{\kappa}_2'=E\left\{(Y-\tilde{\mathbf{x}}'\hat{\boldsymbol{\beta}}_2)\mathbf{h}_0(X, Y)'\right\}\left[E\left\{\mathbf{b}(X)\mathbf{h}_0(X, Y)'\right\}\right]^{-1}$$

and $\mathbf{h}_0(X, Y)=\mathbf{h}(X, Y)\{1-\pi(X, Y)\}$.

Combining (4.5) and (4.7), we obtain

$$\hat{\theta}_{\mathrm{EL}}=\frac{1}{N}\sum_{i=1}^{N}\left\{\hat{y}_i^{(0)}+\frac{\delta_i}{\pi_i}\left(y_i-\hat{y}_i^{(1)}\right)\right\}+o_p(n^{-1/2}), \tag{4.8}$$

where

$$\hat{y}_i^{(0)}=\hat{\beta}_1\{\hat{\pi}(\mathbf{x}_i;\phi)+\mathbf{b}_i'\boldsymbol{\kappa}_1\}+\mathbf{x}_i'\hat{\boldsymbol{\beta}}_2+\mathbf{b}_i'\boldsymbol{\kappa}_2$$

and

$$\hat{y}_i^{(1)} = \hat{\beta}_1 \left(\hat{\pi}_i + \mathbf{b}_i' \boldsymbol{\kappa}_1 \right) + \mathbf{x}_i' \hat{\boldsymbol{\beta}}_2 + \mathbf{b}_i' \boldsymbol{\kappa}_2.$$

The linearization formula (4.8) shows that the EL estimator is asymptotically equivalent to a version of regression estimator but we use $\hat{\pi}(\mathbf{x}_i;\phi)$ instead of $\pi_i = \pi(\mathbf{x}_i; y_i; \phi)$ in $\hat{y}_i^{(0)}$ because y_i are observed only in the sample. The uncertainty associated with $\hat{\boldsymbol{\beta}}$ is asymptotically negligible.

The linearization formula can be used to develop the variance estimation for $\hat{\theta}_{\mathrm{EL}}$. We can use

$$\hat{\boldsymbol{\kappa}}_1' = -\sum_{i=1}^N \hat{\mathbf{g}}_1(\mathbf{x}_i;\hat{\phi})' \left\{ \sum_{i=1}^N \delta_i (\hat{\pi}_i^{-1} - 1) \hat{\mathbf{b}}_i \hat{\mathbf{h}}_i' \right\}^{-1} \tag{4.9}$$

and

$$\hat{\boldsymbol{\kappa}}_2' = \left\{ \sum_{i=1}^N \delta_i \left(\hat{\pi}_i^{-1} - 1 \right) \left(y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_2 \right) \hat{\mathbf{h}}_i' \right\} \left\{ \sum_{i=1}^N \delta_i \left(\hat{\pi}_i^{-1} - 1 \right) \hat{\mathbf{b}}_i \hat{\mathbf{h}}_i' \right\}^{-1} \tag{4.10}$$

to compute the linearization variance estimator.

5. An Illustrative Example

We use a toy example to demonstrate the method and describe the computational details. Suppose that we have two auxiliary variables, X_1 and X_2, and the PS model is given by

$$P(\delta = 1 \mid x_1, x_2, y) = \frac{\exp(\phi_0 + \phi_1 x_1 + \phi_2 y)}{1 + \exp(\phi_0 + \phi_1 x_1 + \phi_2 y)} := \pi(x_1, y; \boldsymbol{\phi}). \tag{5.1}$$

Parameter $\boldsymbol{\phi} = (\phi_0, \phi_1, \phi_2)'$ is estimated by solving

$$\hat{U}_b(\boldsymbol{\phi}) \equiv \frac{1}{N} \sum_{i=1}^N \left\{ \frac{\delta_i}{\pi(x_{1i}, y_i; \boldsymbol{\phi})} - 1 \right\} \mathbf{b}_i = 0 \tag{5.2}$$

where $\mathbf{b}_i = \mathbf{b}(\mathbf{x}_i)$ is a vector such that the solution to (5.2) exists almost everywhere.

The Newton method for solving (5.2) can be written as

$$\hat{\boldsymbol{\phi}}^{(t+1)} = \hat{\boldsymbol{\phi}}^{(t)} + \left\{ \sum_{i=1}^N \delta_i O_i^{(t)} \mathbf{b}_i \mathbf{h}_i' \right\}^{-1} \sum_{i=1}^N \left\{ \frac{\delta_i}{\pi(x_{1i}, y_i; \hat{\boldsymbol{\phi}}^{(t)})} - 1 \right\} \mathbf{b}_i \tag{5.3}$$

where $O_i^{(t)} = \{\pi(x_{1i}, y_i; \hat{\boldsymbol{\phi}}^{(t)})\}^{-1} - 1 = \exp(-\hat{\phi}_0^{(t)} - \hat{\phi}_1^{(t)} x_{1i} - \hat{\phi}_2^{(t)} y_i)$ and $\mathbf{h}_i = (1, x_{1i}, y_i)'$. Because $\left\{ \sum_{i=1}^N \delta_i O_i^{(t)} \mathbf{b}_i \mathbf{h}_i' \right\}$ is not symmetric, the computation for (5.3) can be unstable.

One way to avoid the computational problem is to make the parameter estimation problem an optimization problem. One way is to compute $\hat{\boldsymbol{\phi}}$ by finding the minimizer of

$$Q(\boldsymbol{\phi}) = \hat{U}_b(\boldsymbol{\phi})' \hat{U}_b(\boldsymbol{\phi}).$$

In this case, the Newton method can be expressed by

$$\hat{\boldsymbol{\phi}}^{(t+1)} = \hat{\boldsymbol{\phi}}^{(t)} - \left\{ \dot{U}_b(\hat{\boldsymbol{\phi}}^{(t)})' \dot{U}_b(\hat{\boldsymbol{\phi}}^{(t)}) \right\}^{-1} \dot{U}_b(\hat{\boldsymbol{\phi}}^{(t)})' \hat{U}_b(\hat{\boldsymbol{\phi}}^{(t)}) \tag{5.4}$$

where

$$\dot{U}_b(\hat{\boldsymbol{\phi}}^{(t)})' = -N^{-1} \sum_{i=1}^{N} \delta_i \, O_i^{(t)} \mathbf{b}_i \mathbf{h}_i'.$$

Now, let us discuss how to compute the smoothed propensity score function $\tilde{\pi}(\mathbf{x}; \phi)$ in (3.4). Since $\hat{\boldsymbol{\phi}}$ is obtained by solving (5.2), it satisfies

$$\sum_{i=1}^{N} \delta_i \omega(\mathbf{x}_i, y_i; \hat{\boldsymbol{\phi}}) \mathbf{b}(\mathbf{x}_i) = \sum_{i=1}^{N} \mathbf{b}(\mathbf{x}_i).$$

Thus, by (3.5), the smoothed weight $\widehat{\omega}(\mathbf{x}; \hat{\phi}) = \int \omega(\mathbf{x}, y; \hat{\boldsymbol{\phi}}) \hat{f}_1(y \mid \mathbf{x}) dy$ should also satisfy

$$\sum_{i=1}^{N} \delta_i \widehat{\omega}(\mathbf{x}_i; \hat{\boldsymbol{\phi}}) \mathbf{b}(\mathbf{x}_i) = \sum_{i=1}^{N} \mathbf{b}(\mathbf{x}_i) \tag{5.5}$$

which is the calibration equation for $\mathbf{b}(\mathbf{x})$. That is, the smoothed weights should satisfy the same calibration equation as the original weights.

Since $\omega(\mathbf{x}, y; \boldsymbol{\phi}) = 1 + \exp(-\phi_0 - \phi_1 x_1 - \phi_2 y)$, one way to achieve (5.5) easily is to use

$$\hat{f}_1(y \mid \mathbf{x}) = \frac{1}{\sqrt{2\pi}\hat{\sigma}} \exp\left\{ -\frac{1}{2\hat{\sigma}^2} \left(y - \mathbf{x}'\hat{\boldsymbol{\alpha}} \right)^2 \right\} \tag{5.6}$$

for some $\hat{\boldsymbol{\alpha}}$ and $\hat{\sigma}^2$. We can use the moment-generating function formula of Gaussian distribution to compute

$$\int \exp(-\phi_2 y) \, \hat{f}_1 (y \mid \mathbf{x}) \, dy = \exp\left(-\phi_2 \mathbf{x}'\hat{\boldsymbol{\alpha}} + \frac{1}{2} \phi_2^2 \hat{\sigma}^2 \right). \tag{5.7}$$

We can use constraint (5.5) to compute $\hat{\boldsymbol{\alpha}}$ and $\hat{\sigma}^2$. That is, $\hat{\boldsymbol{\alpha}}$ is the solution to the following estimating equation:

$$\sum_{i=1}^{N} \delta_i \left\{ 1 + \exp\left(-\hat{\phi}_0 - \hat{\phi}_1 x_{1i} - \hat{\phi}_2 \mathbf{x}_i' \boldsymbol{\alpha} + \frac{1}{2} \hat{\phi}_2^2 \hat{\sigma}^2 \right) \right\} \mathbf{b}_i = \sum_{i=1}^{N} \mathbf{b}_i. \tag{5.8}$$

To obtain the unique solution, we may set $\hat{\sigma}^2 = 1$. Once $\hat{\boldsymbol{\alpha}}$ is computed from (5.8), we can obtain

$$\widehat{\pi}(\mathbf{x}_i; \hat{\boldsymbol{\phi}}) = \left\{ 1 + \exp\left(-\hat{\phi}_0 - \hat{\phi}_1 x_{1i} - \hat{\phi}_2 \mathbf{x}_i' \hat{\boldsymbol{\alpha}} + \frac{1}{2} \hat{\phi}_2^2 \hat{\sigma}^2 \right) \right\}^{-1}.$$

Once we obtain $\hat{\pi}(\mathbf{x}; \hat{\boldsymbol{\phi}})$, we can apply the EL method to obtain the final PS weights. The actual computation for the EL weighting can be implemented using the method of Chen et al.[4]. Once $\hat{p}_i$ are obtained by the optimization problem using the bias calibration constraint (4.1) and the balancing constraint (2.2), we can construct the maximum EL estimator of θ by $\hat{\theta}_{\text{EL}} = \sum_{i \in S} \hat{p}_i y_i$.

To compute the linearization variance estimator, we need to compute $\hat{\boldsymbol{\kappa}}_1$ and $\hat{\boldsymbol{\kappa}}_2$ in (4.9) and (4.10), respectively. To compute $\hat{\boldsymbol{\kappa}}_1$, we need to compute $\hat{\mathbf{g}}(\mathbf{x}; \phi)$ in (4.6). Since $\mathbf{h}(\mathbf{x}, y) = (1, x_1, y)'$, we can express

$$\hat{\mathbf{g}}(\mathbf{x}; \phi) = \{\hat{\pi}(\mathbf{x}; \boldsymbol{\phi})\}^2 \left\{ \int \omega(\mathbf{x}, y; \phi)(1, x_1, y)' \hat{f}_1(y \mid \mathbf{x}) dy - 1 \right\} := (\hat{g}_1, \hat{g}_2, \hat{g}_3)'.$$

Since $\hat{\pi}(\mathbf{x}; \phi)$ satisfies (5.5), we can obtain

$$\begin{aligned} \hat{g}_1 &= \{\hat{\pi}(\mathbf{x}; \boldsymbol{\phi})\}^2 \left\{ \int \omega(\mathbf{x}, y; \phi) \hat{f}_1(y \mid \mathbf{x}) dy - 1 \right\} \\ &= \{\hat{\pi}(\mathbf{x}; \boldsymbol{\phi})\}^2 \left\{ \frac{1}{\hat{\pi}(\mathbf{x}; \boldsymbol{\phi})} - 1 \right\} \\ &= \hat{\pi}(\mathbf{x}; \boldsymbol{\phi}) \left\{ 1 - \hat{\pi}(\mathbf{x}; \boldsymbol{\phi}) \right\} \end{aligned}$$

and

$$\hat{g}_2 = x_1 \cdot \hat{\pi}(\mathbf{x}; \boldsymbol{\phi}) \left\{ 1 - \hat{\pi}(\mathbf{x}; \boldsymbol{\phi}) \right\}.$$

Now, to compute $\hat{g}_3$, note that

$$\begin{aligned} \hat{g}_3 &= \{\hat{\pi}(\mathbf{x}; \boldsymbol{\phi})\}^2 \int \exp(-\phi_0 - \phi_1 x_1 - \phi_2 y) y \hat{f}_1(y \mid \mathbf{x}) dy \\ &= \{\hat{\pi}(\mathbf{x}; \boldsymbol{\phi})\}^2 \exp(-\phi_0 - \phi_1 x_1) \int \exp(-\phi_2 y) y \hat{f}_1(y \mid \mathbf{x}) dy. \end{aligned}$$

If $y \mid (\mathbf{x}, \delta = 1) \sim N(\mathbf{x}'\hat{\boldsymbol{\alpha}}, \hat{\sigma}^2)$, then we can obtain

$$\begin{aligned}
\hat{g}_3 &= \{\hat{\pi}(\mathbf{x};\boldsymbol{\phi})\}^2 \exp(-\phi_0 - \phi_1 x_1) \int \exp(-\phi_2 y) y \cdot \frac{1}{\sqrt{2\pi}\hat{\sigma}} \exp\left\{-\frac{1}{2\sigma^2}(y - \mathbf{x}'\hat{\boldsymbol{\alpha}})^2\right\} dy \\
&= \{\hat{\pi}(\mathbf{x};\boldsymbol{\phi})\}^2 \exp(-\phi_0 - \phi_1 x_1) \int y \cdot \frac{1}{\sqrt{2\pi}\hat{\sigma}} \exp\left\{-\frac{1}{2\hat{\sigma}^2}(y - \mathbf{x}'\hat{\boldsymbol{\alpha}} + \hat{\sigma}^2\phi_2)^2\right\} dy \\
&\quad \times \exp\left\{-\frac{1}{\hat{\sigma}^2}(\mathbf{x}'\hat{\boldsymbol{\alpha}})^2 + \frac{1}{2\hat{\sigma}^2}(\mathbf{x}'\hat{\boldsymbol{\alpha}} - \hat{\sigma}^2\phi_2)^2\right\} \\
&= \{\hat{\pi}(\mathbf{x};\boldsymbol{\phi})\}^2 \exp(-\phi_0 - \phi_1 x_1) \exp\left(\frac{1}{2}\phi_2^2\hat{\sigma}^2 - (\mathbf{x}'\hat{\boldsymbol{\alpha}})\phi_2\right) \cdot \left(\mathbf{x}'\hat{\boldsymbol{\alpha}} - \phi_2\hat{\sigma}^2\right) \\
&= \left(\mathbf{x}'\hat{\boldsymbol{\alpha}} - \phi_2\hat{\sigma}^2\right) \cdot \hat{\pi}(\mathbf{x};\boldsymbol{\phi}) \left\{1 - \hat{\pi}(\mathbf{x};\boldsymbol{\phi})\right\},
\end{aligned}$$

where the last equality follows from (5.7).

6. Simulation Study

To evaluate the performance of the proposed methods, we performed a limited simulation study. We consider two outcome models for the simulation study:

1. M1: $y_i = -4 + x_{1i} + x_{2i} + e_i$
2. M2: $y_i = 0.5 \times (x_{1i} + x_{2i} - 5)^2 - 1.5 + e_i$

We use $x_{1i}, x_{2i} \sim N(2, 1)$ and $e_i \sim N(0, 1)$. Regarding the response mechanism, we used $\delta_i \sim$ Bernoulli(π_i) where

$$\pi_i = \frac{\exp(\phi_0 + \phi_1 x_{1i} + \phi_2 y_i)}{1 + \exp(\phi_0 + \phi_1 x_{1i} + \phi_2 y_i)}$$

where $(\phi_0, \phi_1, \phi_2) = (-2, 1.0, 0.5)$. We generated a sample of $(x_{1i}, x_{2i}, y_i, \delta_i)$ from the above mechanism with sample size $N = 5,000$. The overall sampling rate is 50 per cent in both scenarios. We repeat the Monte Carlo sampling independently $B = 1,000$ times.

From each sample, we computed four estimators.

1. (EL-MAR) The EL estimator assuming

$$\pi(\mathbf{x}, y; \phi) = \frac{\exp(\phi_0 + \phi_1 x_{1i} + \phi_2 x_{2i})}{1 + \exp(\phi_0 + \phi_1 x_{1i} + \phi_2 x_{2i})}$$

as the response model. Note that the response model is MAR and incorrectly specified.

Table 1. Monte Carlo Biases, Variance, and Mean Squared Error (MSE) of the Estimators Computed from 1,000 Monte Carlo Samples

Scenario	Method	Bias	Var($\times$1000)	MSE($\times$1000)
M1	Full	0.00	0.55	0.55
	EL (MAR)	0.25	1.34	61.11
	PS	0.00	2.09	2.11
	EL-1	0.00	1.94	1.95
	EL-2	0.01	2.03	2.08
M2	Full	0.00	0.93	0.93
	EL (MAR)	0.62	2.24	386.14
	PS	0.00	2.75	2.75
	EL-1	0.00	2.73	2.73
	EL-2	0.00	2.72	2.72

2. (PS) The PS estimator in (3.9) under the correct model. The parameter $\boldsymbol{\phi}$ for the PS model is estimated by solving

$$\sum_{i=1}^{N}\left(\frac{\delta_i}{\pi(x_{1i}, y_i; \boldsymbol{\phi})} - 1\right)\begin{pmatrix}1\\ \mathbf{x}_i\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}$$

where $\mathbf{x}_i = (x_{1i}, x_{2i})'$.
3. (EL-1) The maximum EL estimator using the estimated π_i in the correct model with the bias calibration condition in (4.1).
4. (EL-2) The maximum EL estimator using estimated π_i under the correct model and satisfying the calibration constraint (2.2) in addition to (4.1).

The simulation results for the point estimators are summarized in Table 1. The performance is as expected. The EL method assuming the MAR model is severely biased. The EL method improves the efficiency over the PS method under the correct model, but the efficiency gain is not high in this simulation setup.

In addition to point estimators, we also calculated the normal-based interval estimator for the proposed EL-2 estimator with 95 per cent nominal coverage rate. The realized coverage rates are 95 per cent and 96 per cent for Scenarios M1 and M2, respectively. Under M2, the linearized variance estimator is slightly overestimated with a relative bias equal to 8 per cent approximately. The slight overestimation of variance for M2 seems to come from the fact that the linear regression model $\hat{f}_1(y \mid \mathbf{x})$ given by (5.6) in computing $\hat{\pi}(\mathbf{x}; \phi)$ through (3.4) is incorrectly specified under M2. Under M1, the normal model assumption for $\hat{f}_1(y \mid \mathbf{x})$ is roughly satisfied and the linearization variance estimator is nearly unbiased.

7. Extension to a Semiparametric PS Model

In this section, we briefly present ideas for an extension to the semiparametric propensity score model. The parametric model approach presented in Section 3 is easy to derive the theory but may be subject to the bias due to model mis-specification. To resolve the problem, we can consider a more flexible propensity score model. One possible model is

$$\pi(\mathbf{x}_i, y_i) \equiv Pr(\delta_i = 1 \mid \mathbf{x}_i, y_i) = \frac{\exp\left\{g(\mathbf{x}_i) + \phi y_i\right\}}{1 + \exp\left\{g(\mathbf{x}_i) + \phi y_i\right\}}, \tag{7.1}$$

where $g(\mathbf{x})$ is completely unspecified. The semiparametric PS model in (7.1) is first considered by Kim and Yu[16] and further discussed by Franks et al.[11]. Model (7.1) implies that

$$f_0(y_i \mid \mathbf{x}_i) = f_1(y_i \mid \mathbf{x}_i) \times \frac{\exp(\gamma y_i)}{E\left\{\exp(\gamma Y) \mid \mathbf{x}_i, \delta_i = 1\right\}}, \tag{7.2}$$

where $\gamma = -\phi$ and $f_\delta(y_i \mid \mathbf{x}_i) = f(y_i \mid \mathbf{x}_i, \delta_i = \delta)$.

Under this model, we can use

$$E\left\{\frac{\delta}{\pi(\mathbf{x}, y)} - 1 \mid \mathbf{x}\right\} = 0$$

to obtain

$$\exp\{g(\mathbf{x})\} = \frac{E\{\delta \exp(\gamma y) \mid \mathbf{x}\}}{E\{1 - \delta \mid \mathbf{x}\}}.$$

For known γ case, we can use kernel regression estimator

$$\exp\{\hat{g}_\gamma(\mathbf{x})\} = \frac{\sum_{i=1}^n \delta_i \exp(\gamma y_i) K_h(x - x_i)}{\sum_{i=1}^n (1 - \delta_i) K_h(x - x_i)}$$

to obtain the following profile PS function

$$\hat{\pi}_p(\mathbf{x}_i, y_i; \gamma) = \frac{\exp\{\hat{g}_\gamma(\mathbf{x}_i) - \gamma y_i\}}{1 + \exp\{\hat{g}_\gamma(\mathbf{x}_i) - \gamma y_i\}}. \tag{7.3}$$

For estimation of γ, Shao and Wang[26] suggested using the GMM method based on some moment conditions. We can use the profile log likelihood to find the maximum likelihood estimator of γ:

$$\ell_p(\gamma) = \sum_{i=1}^N \left[\delta_i \log\{\hat{\pi}_p(\mathbf{x}_i, y_i; \gamma)\} + (1 - \delta_i)\log\{1 - \tilde{\pi}_p(\mathbf{x}_i; \gamma)\}\right] \tag{7.4}$$

where

$$\tilde{\pi}_p(\mathbf{x};\gamma) = \left[\int \{\hat{\pi}_p(\mathbf{x}, y;\gamma)\}^{-1} \hat{f}_1(y \mid \mathbf{x}) dy\right]^{-1}.$$

See Uehara et al.[28] for more details of the profile maximum likelihood estimator of γ. Once $\hat{\gamma}$ is obtained by finding the maximizer of $\ell_p(\gamma)$ in (7.4), we can apply the same EL method to find the final weights. That is, we find the maximizer of

$$\ell(p) = \sum_{i \in S} \log(p_i)$$

subject to $\sum_{i \in S} p_i = 1$,

$$\sum_{i \in S} p_i \hat{\pi}_p(\mathbf{x}_i, y_i; \hat{\gamma}) = N^{-1} \sum_{i=1}^{N} \tilde{\pi}_p(\mathbf{x}_i; \hat{\gamma}),$$

and (2.2). Investigating the asymptotic properties of the resulting EL estimator is beyond the scope of the article and will be presented elsewhere.

8. Concluding Remarks

We have developed an EL-based approach to propensity score estimation with an unknown propensity score function. Assuming that the PS model is correctly specified, we can obtain a consistent estimator of the PS model parameters and construct the final EL weights. The final EL weights can incorporate the benchmarking calibration constraints in addition to the bias correction constraint. The two-step linearization method described in Section 4 can be used to develop a linearized variance estimator of the maximum EL estimator of the population mean. If the PS model is unknown, we may consider a more flexible model, as discussed in Section 7.

There are several possible extensions. Instead of using a single PS model, we can consider multiple PS models and include multiple constraints for bias correction in the EL estimation. Also, the proposed method can be extended to data integration problems (Yang and Kim[31]), combining a voluntary sample with a probability sample. Such extensions will be presented elsewhere.

Appendix

A. Proof of (2.4)

The EL optimization problem can be expressed as maximizing $\ell(\mathbf{p}) = \sum_{i \in S} \log(p_i)$ subject to $\sum_{i \in S} p_i = 1$, bias calibration constraint (2.1), and the benchmarking constraint (2.2). To incorporate the three constraints, the EL weights can be written as

$$\hat{p}_i = \frac{1}{n}\frac{1}{1+\hat{\lambda}_1(\pi_i - W) + \hat{\boldsymbol{\lambda}}_2'(\mathbf{x}_i - \bar{\mathbf{X}}_N)} := \hat{p}_i(\hat{\boldsymbol{\lambda}}),$$

where $\hat{\boldsymbol{\lambda}}' = (\hat{\lambda}_1, \hat{\boldsymbol{\lambda}}_2')$ satisfies $\hat{U}_1(\hat{\boldsymbol{\lambda}}) = 0$ and $\hat{U}_2(\hat{\boldsymbol{\lambda}}) = 0$, and

$$\hat{U}_1(\boldsymbol{\lambda}) = \sum_{i\in S} \hat{p}_i(\boldsymbol{\lambda})\pi_i - W,$$

$$\hat{U}_2(\boldsymbol{\lambda}) = \sum_{i\in S} \hat{p}_i(\boldsymbol{\lambda})\mathbf{x}_i - \bar{\mathbf{X}}_N.$$

Now, under some regularity conditions, we can show that $\hat{\boldsymbol{\lambda}}$ converges in probability to $\boldsymbol{\lambda}^* = (\lambda_1^*, \boldsymbol{\lambda}_2^{*\prime})'$, where $\lambda_1^* = 1/W$ and $\boldsymbol{\lambda}_2^* = \mathbf{0}$. Since we can express $\hat{\theta}_{\mathrm{EL}} = \sum_{i\in S} \hat{p}_i y_i := \hat{\theta}_{\mathrm{EL}}(\hat{\boldsymbol{\lambda}})$ where $\hat{p}_i = \hat{p}_i(\hat{\boldsymbol{\lambda}})$, we can apply Taylor linearization around $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ to get

$$\hat{\theta}_{\mathrm{EL}} = \hat{\theta}_{\mathrm{EL}}(\boldsymbol{\lambda}^*) - \beta_1 \hat{U}_1(\boldsymbol{\lambda}^*) - \boldsymbol{\beta}_2' \hat{U}_2(\boldsymbol{\lambda}^*) + o_p(n^{-1/2}) \tag{8.1}$$

where $(\beta_1, \boldsymbol{\beta}_2')$ satisfies

$$\begin{pmatrix} E\{\frac{\partial}{\partial\lambda_1}\hat{U}_1(\boldsymbol{\lambda}^*)\} & E\{\frac{\partial}{\partial\boldsymbol{\lambda}_2'}\hat{U}_1(\boldsymbol{\lambda}^*)\} \\ E\{\frac{\partial}{\partial\lambda_1}\hat{U}_2(\boldsymbol{\lambda}^*)\} & E\{\frac{\partial}{\partial\boldsymbol{\lambda}_2'}\hat{U}_2(\boldsymbol{\lambda}^*)\} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} = \begin{pmatrix} E\{\frac{\partial}{\partial\lambda_1}\hat{\theta}_{\mathrm{EL}}(\boldsymbol{\lambda}^*)\} \\ E\{\frac{\partial}{\partial\boldsymbol{\lambda}_2}\hat{\theta}_{\mathrm{EL}}(\boldsymbol{\lambda}^*)\} \end{pmatrix}. \tag{8.2}$$

Thus, using $W = N^{-1}\sum_{i=1}^N \pi_i = n/N + O_p(n^{-1/2})$, we can express (8.1) as

$$\begin{aligned} \hat{\theta}_{\mathrm{EL}} &= \frac{W}{n}\sum_{i\in S}\frac{y_i}{\pi_i} + \frac{1}{N}\sum_{i=1}^N\left(\pi_i\beta_1 + \mathbf{x}_i'\boldsymbol{\beta}_2\right) - \frac{W}{n}\sum_{i\in S}\frac{1}{\pi_i}\left(\pi_i\beta_1 + \mathbf{x}_i'\boldsymbol{\beta}_2\right) + o_p(n^{-1/2}) \\ &= \frac{1}{N}\sum_{i=1}^N\left\{\left(\pi_i\beta_1 + \mathbf{x}_i'\boldsymbol{\beta}_2\right) + \frac{\delta_i}{\pi_i}\left(y_i - \pi_i\beta_1 - \mathbf{x}_i'\boldsymbol{\beta}_2\right)\right\} + o_p(n^{-1/2}). \end{aligned} \tag{8.3}$$

Now, to estimate $\boldsymbol{\beta}$, note that (8.2) can be written as

$$E\left\{\sum_{i\in S}\pi_i^{-2}\begin{pmatrix}\pi_i - W \\ \mathbf{x}_i - \bar{\mathbf{X}}_N\end{pmatrix}\begin{pmatrix}\pi_i - W \\ \mathbf{x}_i - \bar{\mathbf{X}}_N\end{pmatrix}'\right\}\begin{pmatrix}\beta_1 \\ \boldsymbol{\beta}_2\end{pmatrix} = E\left\{\sum_{i\in S}\pi_i^{-2}\begin{pmatrix}\pi_i - W \\ \mathbf{x}_i - \bar{\mathbf{X}}_N\end{pmatrix} y_i\right\}.$$

Therefore, (2.4) is proved.

B. Proof of (3.10)

The consistency of $\hat{\phi}$ to ϕ_0 can be obtained by showing $E\{\hat{U}_b(\phi_0)\} = 0$ with some regularity conditions. Let $\hat{\theta}_{\text{PS}}(\hat{\phi}) = N^{-1} \sum_{i \in S} y_i / \pi(\mathbf{x}_i, y_i; \hat{\phi})$. By applying Taylor linearization, we obtain

$$\begin{aligned}
\hat{\theta}_{\text{PS}} &= \hat{\theta}_{\text{PS}}(\phi_0) + \left[E\left\{ \frac{\partial}{\partial \phi'} \hat{\theta}_{\text{PS}}(\phi) \right\} \right] \left(\hat{\phi} - \phi_0 \right) + o_p(n^{-1/2}) \\
&= \frac{1}{N} \sum_{i \in S} \frac{y_i}{\pi_i} - \left[E\left\{ Y \mathbf{h}_0(X, Y)' \right\} \right] \left(\hat{\phi} - \phi_0 \right) + o_p(n^{-1/2}),
\end{aligned}$$

where $\mathbf{b}_i = \mathbf{b}(\mathbf{x}_i; \phi)$ and $\mathbf{h}_0(\mathbf{x}, y; \phi) = \mathbf{h}(\mathbf{x}, y; \phi)\{1 - \pi(\mathbf{x}, y; \phi)\}$.

Also, by Taylor linearization of $\hat{U}_b(\hat{\phi}) = 0$, we obtain

$$\begin{aligned}
\hat{\phi} - \phi_0 &= -\left[E\left\{ \frac{\partial}{\partial \phi'} \hat{U}_b(\phi_0) \right\} \right]^{-1} \hat{U}_b(\phi_0) + o_p(n^{-1/2}) \\
&= \left[E\left\{ \mathbf{b}(X) \mathbf{h}_0(X, Y)' \right\} \right]^{-1} \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{\delta_i}{\pi(\mathbf{x}_i, y_i; \phi)} - 1 \right\} \mathbf{b}_i + o_p(n^{-1/2}).
\end{aligned}$$

Thus, combining the two results, we can obtain

$$\hat{\theta}_{\text{PS}} = \frac{1}{N} \sum_{i \in S} \frac{y_i}{\pi_i} - \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{\delta_i}{\pi(\mathbf{x}_i, y_i; \phi)} - 1 \right\} \boldsymbol{\gamma}' \mathbf{b}_i + o_p(n^{-1/2}),$$

where

$$\boldsymbol{\gamma}' = \left[E\left\{ Y \mathbf{h}_0(X, Y)' \right\} \right] \left[E\left\{ \mathbf{b}(X) \mathbf{h}_0(X, Y)' \right\} \right]^{-1}.$$

Therefore, (3.10) is established.

Acknowledgements

The article has some overlap with the contents in the paper of Wang, Shao, and Kim (2014) published in Statistica Sinica (doi:10.5705/ss.2012.074), particularly in the introduction section. The authors thank the guest editor, Professor Partha Lahiri, for invitation to the special issue and for the constructive comments.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Research by the first author was partially supported by a grant from the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. Research by the second author was supported by JSPS Grant-in-Aid for Scientific Research(A) 23H0046 and MEXT Project for Seismology toward Research Innovation with Data of Earthquake (STAR-E) Grant Number JPJ010217.

References

1. Baker R, Brick JM, Bates NA, Battaglia M, Couper MP, Dever JA, Gile KJ and Tourangeau R. Summary report of the AAPOR task force on non-probability sampling. *J Sur Stat Meth* 2013; 1: 90–143.
2. Berger YG and Torres ODLR. An empirical likelihood approach for inference under complex sampling design. *J Royal Stat Soc, Series B* 2016; 78: 319–341.
3. Cao W, Tsiatis AA and Davidian M. Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika* 2009; 96: 723–734.
4. Chen J, Sitter R and Wu C. Using empirical likelihood method to obtain range restricted weights in regression estimators for surveys. *Biometrika* 2002; 89: 230–237.
5. Cochran WG. *Sampling techniques* (3rd ed.). John Wiley & Sons 1977.
6. Copas JB and Eguchi S. Local sensitivity approximations for selectivity bias. *J Royal Stat Soc: Series B* 2001; 63: 871–895.
7. Copas JB and Li HG. Inference for non-random samples. *J Royal Stat Soc: Series B* 1997; 59: 55–95.
8. Dever JA and Valliant R. General regression estimation adjusted for undercoverage and estimated control totals. *J Surv Stat Meth* 2016; 4: 289–318.
9. Elliott M and Valliant R. Inference for nonprobability samples. *Stat Sci* 2017; 32: 249–264.
10. Firth D and Bennett KE. Robust models in probability sampling. *J Royal Stat Soc, Series B* 1998; 60: 3–21.
11. Franks AM, Airoldi EM and Rubin DB. Nonstandard conditionally specified models for nonignorable missing data. *Proc National Acad of Sci* 2022; 117: 19045–19053.
12. Fuller WA. *Sampling statistics*. Hoboken, NJ: John Wiley & Sons, Inc 2009.
13. Kim JK. Calibration estimation using empirical likelihood in survey sampling. *Stat Sinica* 2009; 19: 145–158.
14. Kim JK and Shao J. *Statistical methods for handling incomplete data* (2nd ed.). CRC press 2021.
15. Kim JK and Skinner CJ. Weighting in survey analysis under informative sampling. *Biometrika* 2013; 100: 358–398.
16. Kim JK and Yu CL. A semi-parametric estimation of mean functionals with non-ignorable missing data. *J Amer Stat Assoc* 2011; 106: 157–165.
17. Kott PS and Chang T. Using calibration weighting to adjust for nonignorable unit nonresponse. *J Amer Stat Assoc* 2010; 105: 1265–1275.
18. Morikawa K and Kim JK. Semiparametric optimal estimation with nonignorable nonresponse data. *Ann Stat* 2021; 49: 2991–3014.
19. Morikawa K, Kim JK and Kano Y. Semiparametric maximum likelihood estimation under nonignorable nonresponse. *Canadian J Stat* 2017; 45: 393–409.
20. Pfeffermann D and Sikov A. Imputation and estimation under nonignorable nonresponse in household surveys with missing covariate information. *J Off Stat* 2011; 27: 181–209.
21. Pfeffermann D and Sverchkov M. Parametric and semiparametric estimation of regression models fitted to survey data. *Sankhyā, Series B* 1999; 61: 166–186.

22. Qin J, Leung D and Shao J. Estimation with survey data under non-ignorable nonresponse or informative sampling. *J Amer Stat Assoc* 2002; 97: 193–200.
23. Riddles MK, Kim JK and Im J. Propensity-score-adjustment method for nonignorable nonresponse. *J Surv Stat Meth* 2016; 4: 215–245.
24. Rubin DB. Inference and missing data. *Biometrika* 1976; 63: 581–592.
25. Särndal CE, Cassel CM and Wretman JH. *Model assisted survey sampling*. New York: Springer-Verlag 1992.
26. Shao J and Wang L. Semiparametric inverse propensity weighting for nonignorable missing data. *Biometrika* 2016; 103: 175–187.
27. Tillé Y. *Sampling and estimation from finite populations*. John Wiley & Sons 2020.
28. Uehara M, Lee D and Kim JK. Semiparametric response model with nonignorable nonresponse. *Scandinavian J Stat* 2023. doi: 10.1111/sjos.12652.
29. Wang S, Shao J and Kim JK. An instrument variable approach for identification and estimation with nonignorable nonresponse. *Stat Sinica* 2014; 24: 1097–1116.
30. Wu C and Sitter RR. A model-calibration approach to using complete auxiliary information from survey data. *J Amer Stat Assoc* 2001; 96: 185–193.
31. Yang S and Kim JK. Statistical data integration in survey sampling: A review. *Japanese J Stat Data Sci* 2020; 3: 625–650.