Statistics and Probability Letters 140 (2018) 1–6

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A note on the equivalence of two semiparametric estimation methods for nonignorable nonresponse

Kosuke Morikawa [a,*], Jae Kwang Kim [b,c]

[a] Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan
[b] Department of Statistics, Iowa State University, Ames, IA 50011, USA
[c] Department of Mathematical Sciences, KAIST, Daejeon, 34141, Republic of Korea

ARTICLE INFO

Article history:
Received 30 October 2017
Received in revised form 20 March 2018
Accepted 31 March 2018
Available online 24 April 2018

MSC:
62D99
62F12

Keywords:
Empirical likelihood
Incomplete data
Nonignorable
Nonresponse
Not missing at random (NMAR)
Semiparametric estimation

ABSTRACT

We consider semiparametric estimation with nonignorable nonresponse data where only a parametric response model is assumed. We clarify the relationship of existing estimators and propose a new estimator which attains the semiparametric efficiency bound and is robust to model misspecification.

1. Introduction

Missing data problems are ubiquitous in many research areas, including econometrics, epidemiology, clinical study, and psychometrics. If analysts do not properly deal with missing data, then the results may be biased, which can lead to incorrect conclusions. Thus, a proper method for analyzing missing data needs to be developed. Also, it is preferable that the required assumptions in the proposed method be as weak as possible.

The required assumptions are strongly related to the outcome model or the response mechanism. In this paper, we focus on estimation with nonresponse data in which study variable is subject to missingness. Let y be the study variable, $\boldsymbol{x}$ be a fully observed d-dimensional covariate vector, and r be a response indicator of y, i.e., r takes the value 1 if y is observed, and takes the value 0 if y is missing. Thus, letting $\boldsymbol{z} = (\boldsymbol{x}^\top, y)^\top$, we observe $(\boldsymbol{x}, y)$ when $r = 1$, and observe only $\boldsymbol{x}$ when $r = 0$. The response mechanism is defined as the conditional probability $\pi(\boldsymbol{z}) = \Pr(R = 1 \mid \boldsymbol{z})$. If the mechanism does not depend on the study variable y, then it is called missing at random (MAR) and otherwise is called not missing at random (NMAR) (Little and Rubin, 2002; Kim and Shao, 2013). In the analysis of nonresponse data, MAR (NMAR) is also referred to as ignorable (nonignorable) missingness.

In this paper, we assume a parametric model on the response mechanism. Let $\pi(\boldsymbol{z}; \boldsymbol{\phi})$ be the parametric response model, where $\boldsymbol{\phi}$ is a q-dimensional parameter. In classical approaches for NMAR data, an outcome model $f(y \mid \boldsymbol{x})$, which is the

* Corresponding author.
E-mail addresses: morikawa@sigmath.es.osaka-u.ac.jp (K. Morikawa), jkim@iastate.edu (J.K. Kim).

https://doi.org/10.1016/j.spl.2018.03.020
0167-7152/

conditional distribution of y given $\boldsymbol{x}$, is assumed in addition to the response model (Greenlees et al., 1982). This estimator has been criticized because of its sensitivity to model assumptions. Recently, some semiparametric methods, which do not require any outcome model, have been proposed.

Semiparametric estimation is mainly divided into two approaches: (i) the empirical likelihood (EL) approach; and (ii) the moment-based approach. Qin et al. (2002) derived a consistent and asymptotic normal estimator for $\boldsymbol{\phi}$ by using a technique of EL without using any outcome model. Kott and Chang (2010) proposed a moment-based estimator for $\boldsymbol{\phi}$, also without using any outcome model. Recently, Morikawa and Kim (2016) proposed two moment-based semiparametric adaptive estimators.

In this paper, we clarify the relationship between the EL estimator and the moment-base estimator, and show that there exists a specific case for which these two estimators are exactly the same. Also, we propose an estimation method that is robust to model misspecification. All technical details are given in the Supplementary Material (see Appendix A).

2. Previous semiparametric estimators

Let $\boldsymbol{z}_i = (\boldsymbol{x}_i, y_i)^\top$ $(i = 1, \ldots, n)$ be independently and identically distributed realizations from unknown distribution $F(\boldsymbol{z})$, and r_i $(i = 1, \ldots, n)$ be independently distributed taking binary values, either 0 or 1, with probability $\Pr(R_i = 1 \mid \boldsymbol{z}_i) = \pi(\boldsymbol{z}_i)$ for $i = 1, \ldots, n$. Also, without loss of generality, assume that the first m elements are observed and that the remaining $(n - m)$ elements are missing in y_i, i.e., $r_i = 1$ for $i = 1, \ldots, m$ and $r_i = 0$ for $i = m + 1, \ldots, n$. Qin et al. (2002) constructed the likelihood without using the data when $r = 0$ by

$$\prod_{i=1}^{m} \pi(\boldsymbol{\phi}; \boldsymbol{z}_i) dF(\boldsymbol{z}_i) \prod_{i=m+1}^{n} \int \{1 - \pi(\boldsymbol{\phi}; \boldsymbol{z})\} dF(\boldsymbol{z}) \tag{1}$$

and discretized the distribution F by w_i $(i = 1, \ldots, m)$. The discretized distribution w_i can be estimated by maximizing $\prod_{i=1}^{m} w_i$ under the following constraints:

$$w_i \geq 0, \quad \sum_{i=1}^{m} w_i = 1, \quad \sum_{i=1}^{m} w_i \{\pi(\boldsymbol{\phi}; \boldsymbol{z}_i) - W\} = 0,$$

$W = \Pr(R = 1) = \int \pi(\boldsymbol{z}; \boldsymbol{\phi}_0) dF(\boldsymbol{z})$, and $\sum_{i=1}^{m} w_i \{\boldsymbol{h}(\boldsymbol{x}_i) - \bar{\boldsymbol{h}}_n\} = 0$, where $\boldsymbol{h} : \mathbb{R}^d \to \mathbb{R}^{p_1}$ $(p_1 \geq q-1)$ is an arbitrary function of $\boldsymbol{x}$, and $\bar{\boldsymbol{h}}_n = n^{-1} \sum_{i=1}^{n} \boldsymbol{h}(\boldsymbol{x}_i)$. The $\boldsymbol{h}(\boldsymbol{x})$ function helps to improve the efficiency. By introducing Lagrange multipliers, the solution to the above optimization problem is $\hat{w}_i^{-1} = m[1 + \boldsymbol{\lambda}_1^\top \{\boldsymbol{h}(\boldsymbol{x}_i) - \bar{\boldsymbol{h}}_n\} + \lambda_2 \{\pi(\boldsymbol{\phi}; \boldsymbol{z}_i) - W\}]$. By profiling out the unknown F with the estimates $\hat{w}_i$ $(i = 1, \ldots, m)$ in (1) and taking the logarithm, we obtain the profile pseudo-loglikelihood:

$$\begin{aligned} &\ell(\boldsymbol{\phi}, W, \boldsymbol{\lambda}_1) \\ &= \sum_{i=1}^{m} \log \pi(\boldsymbol{\phi}; \boldsymbol{z}_i) - \sum_{i=1}^{m} \log[1 + \boldsymbol{\lambda}_1^\top \{\boldsymbol{h}(\boldsymbol{x}_i) - \bar{\boldsymbol{h}}_n\} + \lambda_2 \{\pi(\boldsymbol{\phi}; \boldsymbol{z}_i) - W\}] \\ &\quad + (n - m) \log(1 - W), \end{aligned} \tag{2}$$

where $\lambda_2 = (n/m - 1)/(1 - W)$. Qin et al. (2002) proposed a semiparametric estimator for $\boldsymbol{\phi}$ that maximizes the profile pseudo-loglikelihood. In the optimization procedure, some computational techniques are needed (see Chen et al., 2002) because the maximizer of (2) must satisfy $\hat{w}_i \geq 0$.

On the other hand, under the same assumptions, Kott and Chang (2010) proposed another semiparametric estimator that solves the following estimating equation:

$$\sum_{i=1}^{n} \left\{ \frac{r_i}{\pi(\boldsymbol{\phi}; \boldsymbol{z}_i)} - 1 \right\} \boldsymbol{g}(\boldsymbol{x}_i) = 0, \tag{3}$$

where $\boldsymbol{g} : \mathbb{R}^d \to \mathbb{R}^q$ is an arbitrary function of $\boldsymbol{x}$. This equation is called "calibration" equation in the literature of survey sampling. A typical choice for $\boldsymbol{g}$ when $d = 1$ is $\boldsymbol{g}(x) = (1, x, \ldots, x^{q-1})^\top$. It is hard to decide the control variables in the calibration condition when $d > 1$. Also, when the dimension of $\boldsymbol{g}(\boldsymbol{x})$ is larger than q, say p_2, the model is over-identified and the generalized method of moments (GMM) method (Hansen, 1982) can be used to estimate $\boldsymbol{\phi}$. The GMM estimator can be constructed by

$$\hat{\boldsymbol{\phi}} := \arg\min_{\boldsymbol{\phi}} \sum_{i=1}^{n} \left\{ \frac{r_i}{\pi(\boldsymbol{\phi}; \boldsymbol{z}_i)} - 1 \right\}^2 \boldsymbol{g}(\boldsymbol{x}_i)^\top \hat{V}^{-1}(\boldsymbol{\phi}) \boldsymbol{g}(\boldsymbol{x}_i), \tag{4}$$

where $\hat{V}(\boldsymbol{\phi}) = n^{-1} \sum_{i=1}^{n} \{r_i / \pi(\boldsymbol{\phi}; \boldsymbol{z}_i) - 1\}^2 g(\boldsymbol{x}_i)^{\otimes 2}$ and $B^{\otimes 2} = BB^\top$ for any matrix B. The optimizations in (3) and (4) are much simpler than that of Qin et al. (2002) since there is no constraint in the optimization.

Remark 2.1. Regarding the choice of $\boldsymbol{g}(\boldsymbol{x})$ in (3), Morikawa and Kim (2016) proposed an optimal $\boldsymbol{g}$ function,

$$\boldsymbol{g}^{\star}(\boldsymbol{x}) = \frac{\mathrm{E}_1\{\dot{\pi}(\boldsymbol{\phi}_0; \boldsymbol{x}, Y)/\pi(\boldsymbol{\phi}_0; \boldsymbol{x}, Y)^2 \mid \boldsymbol{x}\}}{\mathrm{E}_1[\{1-\pi(\boldsymbol{\phi}_0; \boldsymbol{x}, Y)\}/\pi(\boldsymbol{\phi}_0; \boldsymbol{x}, Y)^2 \mid \boldsymbol{x}]}, \tag{5}$$

where $\mathrm{E}_1(\cdot \mid \boldsymbol{x})$ is the conditional expectation on y given $\boldsymbol{x}$ and $r = 1$. Let $f_1(y \mid \boldsymbol{x}; \boldsymbol{\beta})$ be the working model of the conditional distribution of y given $\boldsymbol{x}$ and $r = 1$, and $\hat{\boldsymbol{\beta}}$ be the maximum likelihood estimator of $\boldsymbol{\beta}$ obtained from the set of full respondents. Then it can be shown that the solution of (3) with $\boldsymbol{g}^{\star}(\boldsymbol{x})$ and $f_1(y \mid \boldsymbol{x}; \hat{\boldsymbol{\beta}})$ attains the semiparametric efficiency bound when the f_1 model is correctly specified. Also even if the f_1 model is wrong, the estimator has consistency and asymptotic normality.

Furthermore, nonparametric estimation can be used for both of denominator and numerator in (5). For example, the denominator can be estimated by using a kernel smoothing without assuming any parametric model on f_1 by

$$\frac{1}{n}\sum_{i=1}^{n} r_i K\left(\frac{\boldsymbol{x}_i - \boldsymbol{x}}{h}\right)\frac{1-\pi(\boldsymbol{\phi}_0; \boldsymbol{x}_i, y_i)}{\pi(\boldsymbol{\phi}_0; \boldsymbol{x}_i, y_i)^2},$$

where $K(\cdot)$ is a kernel function and h is a bandwidth satisfying some regularity conditions, see Morikawa and Kim (2016) for more details.

The two semiparametric estimation methods use the same assumptions but they seem to provide different estimation results. One natural question is which one is better, or whether there are any real differences between the two methods. We will answer these questions in the next section.

3. Theoretical comparison of the semiparametric estimators

In this section, we show that when $q = p_1 + 1 = p_2$, for a specific choice of $\boldsymbol{g}(\boldsymbol{x})$ and $\boldsymbol{h}(\boldsymbol{x})$ functions, the two estimators are exactly the same.

Theorem 3.1. *When $q = p_1 + 1 = p_2$, EL and GMM estimators are exactly the same if and only if $\boldsymbol{g}(\boldsymbol{x}) = \{1, \boldsymbol{h}(\boldsymbol{x})^{\top}\}^{\top}$.*

We now consider the other case of estimating $\boldsymbol{\phi}$ by using the method of Qin et al. (2002) with the same $\boldsymbol{h}(\boldsymbol{x})$ function. Theorem 3.1 implies that, when $q = p_1 + 1$, there is no reason to use the procedure for EL, so we just use the GMM estimating Eq. (3) with $\boldsymbol{g}(\boldsymbol{x}) = \{1, \boldsymbol{h}(\boldsymbol{x})^{\top}\}^{\top}$, which is much simpler in terms of computation.

Furthermore, we can see that the GMM estimator including $\boldsymbol{g}^{\star}(\boldsymbol{x})$ as the constraint also attains the semiparametric efficiency bound.

Theorem 3.2. *When $q < p_1 + 1 = p_2$, the GMM estimator attains the semiparametric efficiency bound if $\boldsymbol{g}(\boldsymbol{x})$ defined in (4) contains $\boldsymbol{g}^{\star}(\boldsymbol{x})$, i.e., $\boldsymbol{g}(\boldsymbol{x}) = \{\boldsymbol{k}(\boldsymbol{x})^{\top}, \boldsymbol{g}^{\star}(\boldsymbol{x})^{\top}\}^{\top}$, where $\boldsymbol{k} : \mathbb{R}^d \to \mathbb{R}^{\kappa}$ is an arbitrary function of $\boldsymbol{x}$, $\kappa \in \mathbb{R}^1$ is a positive integer, and $\boldsymbol{g}^{\star}(\boldsymbol{x})$ is defined in Remark 2.1.*

This estimator enjoys two properties: (i) robustness to misspecification of the response model; and (ii) semiparametric efficiency. The robustness is achieved because of the constraints, which are proposed by Qin and Zhang (2007) and Imai and Ratkovic (2014) for observational studies in the case of MAR. Our study can be considered as an extension of their work to NMAR.

4. Simulation study

We conduct a simulation study (i) to confirm the asymptotic behavior of EL and GMM estimators when $q < p_1 + 1 = p_2$ and $\boldsymbol{g}(\boldsymbol{x}) = \{1, \boldsymbol{h}(\boldsymbol{x})^{\top}\}^{\top}$; and (ii) to check the performance of the estimators proposed in Theorem 3.2 when both the response and $f_1(y \mid x)$ models are misspecified. In Theorem 3.1, we have shown that EL and GMM estimators are exactly the same when $q = p_1 + 1 = p_2$. However, we have not shown anything about the relationship between the two types of estimators when $q < p_1 + 1 = p_2$. Thus, we check the relationship here. For (ii), in Theorem 3.2, we have shown that any function $\boldsymbol{k}(\boldsymbol{x})$ can be included in our GMM constraints without loss of efficiency as long as $\boldsymbol{g}^{*}(\boldsymbol{x})$ is used. We check that the other function $\boldsymbol{k}(\boldsymbol{x})$ can potentially reduce the bias when both models are misspecified.

Let the covariate be $X \sim \mathcal{N}(-1, 1)$ and $Y \mid X = x \sim \mathcal{N}(2x(x-3/4)(x+3/4), 1/3)$. We prepare six different response mechanisms: M1 (linear nonignorable): $\mathrm{logit}\{\pi(y)\} = 0.90 - 1.0y$; M2 (linear ignorable): $\mathrm{logit}\{\pi(x)\} = 0.90 - 1.0x$; M3 (quadratic nonignorable): $\mathrm{logit}\{\pi(y)\} = 1.3 - 0.5y - y^2$; and M4 (log–log nonignorable): $\pi(x, y) = 1 - \exp\{-\exp(0.25 + 0.5x - y)\}$. In this simulation study, we specify a response model by $\pi(y; \boldsymbol{\phi}) = \{1 + \exp(\phi_0 + \phi_1 y)\}^{-1}$, where $\boldsymbol{\phi} = (\phi_0, \phi_1)^{\top}$. Therefore, all models except for M1 are misspecified.

Furthermore, we assume that $f_1(y \mid x)$ is normally distributed with a mean function $\mu(x; \boldsymbol{\beta})$ and variance σ^2, although, strictly speaking, this $f_1(y \mid x)$ misspecifies the true f_1, because it is given by $f_1(y \mid x) \propto \pi(x, y)f(y \mid x)$. Thus, the true model for $f_1(y \mid x)$ depends on its response mechanism as well as the distribution of outcome, and is not normal. However, as is stated in Remark 2.1, this misspecification does not affect the consistency of $\hat{\boldsymbol{\phi}}$. We consider two settings for the mean function: (1) (Heavily misspecified model) $\mu_1(x; \boldsymbol{\beta}) = \beta_0 + \beta_1 x$; (2) (Slightly misspecified model) $\mu_2(x; \boldsymbol{\beta}) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$, where in the second model an appropriate $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top$ is chosen by the Akaike information criterion (AIC). The first mean function is linear with respect to x, whereas the second function is cubic. It is expected that estimators with the second mean function are more efficient than those with the first one.

For EL estimators, we consider three types of $\boldsymbol{h}(\boldsymbol{x})$ function: (1) $\boldsymbol{h}_1(x) = (x, x^2)^\top$; (2) $\boldsymbol{h}_2(x) = \boldsymbol{g}^\star(x)$; and (3) $\boldsymbol{h}_3(x) = \{x, \boldsymbol{g}^\star(x)^\top\}^\top$. For GMM estimators, we consider four types of $\boldsymbol{g}(\boldsymbol{x})$ function: (1) $\boldsymbol{g}_1(x) = (1, x, x^2)^\top$; (2) $\boldsymbol{g}_2(x) = \boldsymbol{g}^\star(x)$; (3) $\boldsymbol{g}_3(x) = \{1, \boldsymbol{g}^\star(x)^\top\}^\top$; and (4) $\boldsymbol{g}_4(x) = \{1, x, \boldsymbol{g}^\star(x)^\top\}^\top$. We can check (i) the asymptotic behaviors of EL and GMM estimators when $q < p_1 + 1 = p_2$ and $\boldsymbol{g}(\boldsymbol{x}) = \{1, \boldsymbol{h}(x)^\top\}^\top$, for example, by comparing estimators using $\boldsymbol{h}_1(x)$ and $\boldsymbol{g}_1(x)$, because $\boldsymbol{g}_1(x) = \{1, \boldsymbol{h}_1(x)^\top\}^\top$. The parameter $\boldsymbol{\phi}$ is estimated by (3) when $q = p_1 + 1 = p_2$, and by (4) when $q < p_1 + 1 = p_2$, for which two-step GMM estimation procedure is used. Instead of using the two-step procedure, there are some different versions of GMM procedures such as continuous updating procedure (Pakes and Pollard, 1989; Hansen et al., 1996). Although the small size property of the procedures is somewhat different, the asymptotic properties are the same. We use the two-step procedure because the sample size is large enough and the procedure is easy to conduct. The algorithm can be easily implemented by using the package "gmm" in the R programing language.

It is meaningless to compare the estimated parameters of the response model because it is misspecified in models M2–M4, and the true parameter value is unknown. Therefore, we assess robustness by estimating $\theta = E(Y)$ with the estimated response model, as follows:

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{r_i y_i}{\pi(y_i; \hat{\boldsymbol{\phi}})} + \left\{1 - \frac{r_i}{\pi(y_i; \hat{\boldsymbol{\phi}})}\right\} \mathrm{E}^\star(Y \mid x_i; \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\beta}})\right],$$

where

$$\mathrm{E}^\star(Y \mid x; \boldsymbol{\phi}, \boldsymbol{\beta}) = \frac{\mathrm{E}_1[Y\{1 - \pi(Y; \boldsymbol{\phi})\}/\pi(Y; \boldsymbol{\phi})^2 \mid x]}{\mathrm{E}_1[\{1 - \pi(Y; \boldsymbol{\phi})\}/\pi(Y; \boldsymbol{\phi})^2 \mid x]}.$$

This estimator attains the semiparametric efficiency bound when both $\pi(y; \boldsymbol{\phi})$ and $f_1(y \mid x; \boldsymbol{\beta})$ are correctly specified; however, it still has consistency even if $f_1(y \mid x; \boldsymbol{\beta})$ is misspecified (Morikawa and Kim, 2016).

Under this setup, Monte Carlo samples of size $n = 2000$ are generated independently with 2,000 replications. The results with response mechanisms M1–M4 are shown in Fig. 1. In the boxplots, the samples not having converged are removed. In particular, for EL estimators with $\boldsymbol{h}_3(x)$ under the heavily misspecified mean function $\mu_1(x)$, almost all samples did not converge for all response mechanisms while $\boldsymbol{g}_4(x)$ with the same mean structure did, which indicates the difficulty of optimization in the EL method. For estimators using $\boldsymbol{g}_2(x)$ and $\boldsymbol{h}_2(x)$ with $\mu_1(x)$, nearly half of the samples did not converge. For the other estimators, most of samples converged except for those using $\boldsymbol{g}_4(x)$ with $\mu_1(x)$ in M3. This result tells us importance of the correct specification of the model for $f_1(y \mid x; \beta)$. Furthermore, all estimators are correctly estimated in M1, i.e., when the response model is correct. However, even when the response model is misspecified, the estimators using $\mu_2(x)$ are quite well estimated. Overall, the performance of the GMM estimator using $\boldsymbol{g}_2(x)$ or $\boldsymbol{g}_3(x)$ with mean function $\mu_2(x)$ is the best. Surprisingly, the estimators in M2 still work well even though the true response mechanism is MAR.

We would like to note the following about issue (i). It can be inferred from the results of estimators using $\boldsymbol{g}_1(x)$ and $\boldsymbol{h}_1(x)$ under mechanisms M1–M4 that Theorem 3.1 does not hold when $q < p_1 + 1 = p_2$. This implies that EL estimators may not attain the semiparametric efficiency bound. In the case of (ii), we can see that having more constraints leads to less biased estimators by comparing estimators using $\boldsymbol{g}_2(x)$, $\boldsymbol{g}_3(x)$, and $\boldsymbol{g}_4(x)$ with $\mu_1(x)$ as the mean function, though some estimators using the most constrained function $\boldsymbol{g}_4(x)$ were not estimated well. However, on the flip side, it follows from the results of estimators using $\boldsymbol{g}_4(x)$ with $\mu_2(x)$ as the mean function that having more constraints leads to difficulty in the optimization computation or a lack of identification. Therefore, choosing an appropriate constraint function $\boldsymbol{k}(x)$ in $\boldsymbol{g}(x) = \{\boldsymbol{g}^\star(x)^\top, \boldsymbol{k}(x)^\top\}^\top$ from observed data is important. One way to check the over-identification of the $\boldsymbol{k}(x)$ function is to use Sargan–Hansen's J test (Hansen, 1982). We have conducted the Sargan–Hansen's J test, which is shown in the Supplementary Material (see Appendix A).

It is of interest to know ignorability of the response model because if the response mechanism is MAR, then we do not have to model the outcome model at all or can use estimators enjoying fascinating properties such as the double or multiple robustness (Robins et al., 1994; Han, 2014). The result is shown in the Supplementary Material (see Appendix A).

5. Conclusion

In this paper, we have investigated the relationship between the two semiparametric methods, the EL and GMM approaches, for estimation of the response model. We have shown that the EL estimators can be solved by using the moment-based method when $q = p_1 + 1 = p_2$, where p_1 and p_2 are defined in Section 2. Also, we have pointed out that EL and GMM estimators may not be asymptotically equivalent when $q < p_1 + 1 = p_2$, based on a numerical study. Furthermore, we have shown that a constraint function including the best function $g^\star(\boldsymbol{x})$, defined in Remark 2.1, also attains the semiparametric efficiency bound (Rotnitzky and Robins, 1997; Morikawa and Kim, 2016). This property is useful in practice because the estimated response model is robust to model misspecification due to the other constraints, and the true response mechanism is generally unknown.

However, as is shown in Section 4, having more constraints may lead to a lack of identification which leads to numerical problems. Thus, a method for choosing an appropriate constraint function will be a topic of our future research.

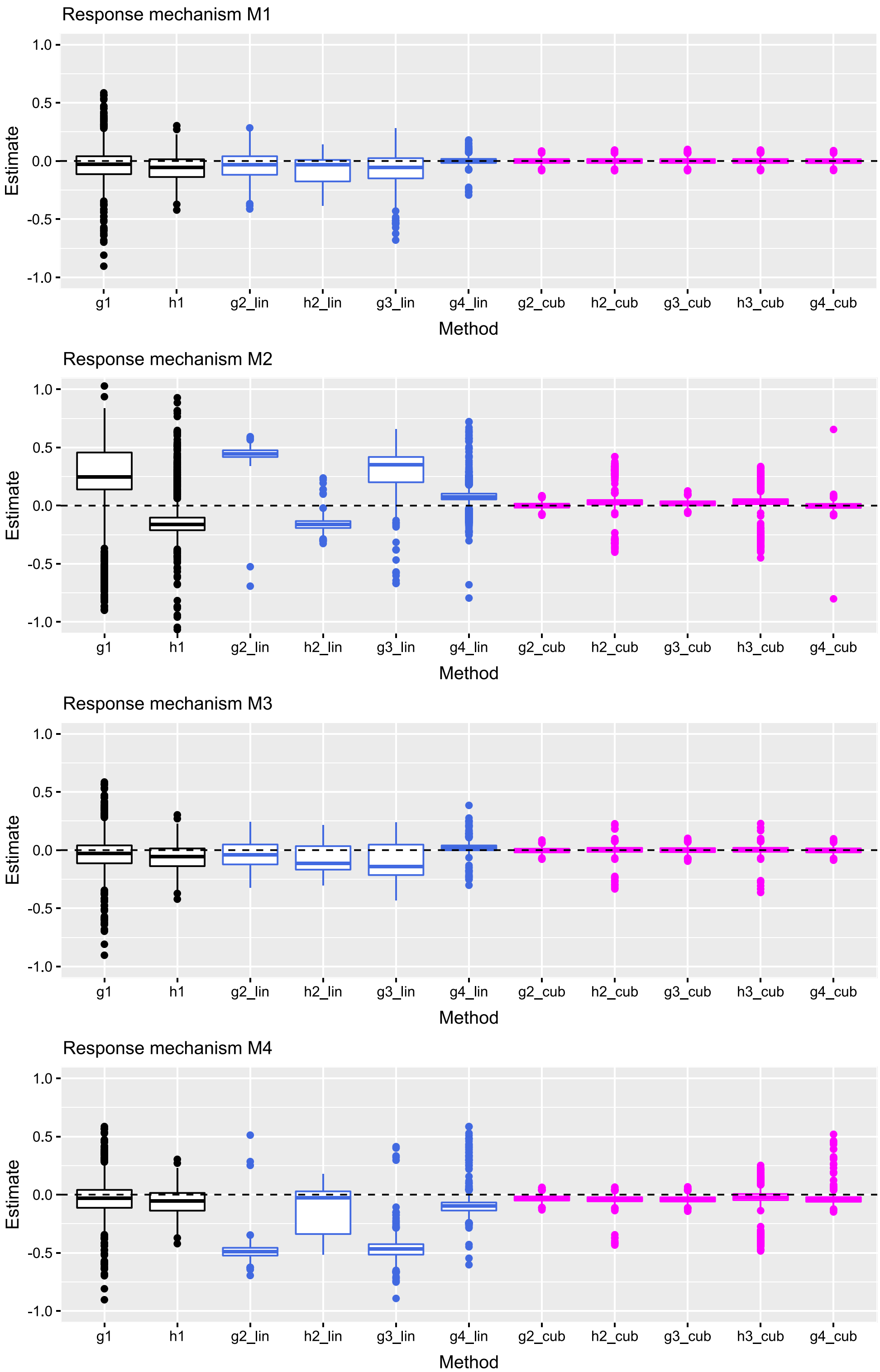

Fig. 1. Boxplots of eleven estimators for $E(Y)$ with M1–M4 response mechanisms. Black indicates without the $f_1(y \mid x)$ model, blue indicates with mean function $\mu_1(x)$ (heavily misspecified model), and pink indicates with $\mu_2(x)$ (slightly misspecified model). The first element of the method name indicates the constraint used for the estimators, and the second element, "lin" or "cub", refers to either the linear or cubic model, respectively, for the mean function of $f_1(y \mid x)$.

Acknowledgments

The research of the first author was partially supported by JSPS KAKENHI Grant Number JP16J03875. The research of the second author was partially supported by the U.S. National Science Foundation (Grant Number #1733572).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2018.03.020.

References

Chen, J., Sitter, R.R., Wu, C., 2002. Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. Biometrika 89, 230–237.
Greenlees, J.S., Reece, W.S., Zieschang, K.D., 1982. Imputation of missing values when the probability of response depends on the variable being imputed. J. Amer. Statist. Assoc. 77, 251–261.
Han, P., 2014. Multiply robust estimation in regression analysis with missing data. J. Amer. Statist. Assoc. 109, 1159–1173.
Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.
Hansen, L.P., Heaton, H., Yaron, A., 1996. Finite-sample properties of some alternative gmm estimators. J. Bus. Econom. Statist. 14, 262–280.
Imai, K., Ratkovic, M., 2014. Covariate balancing propensity score. J. R. Stat. Soc. Ser. B Stat. Methodol. 76, 243–263.
Kim, J.K., Shao, J., 2013. Statistical Methods for Handling Incomplete Data. CRC Press.
Kott, P.S., Chang, T., 2010. Using calibration weighting to adjust for nonignorable unit nonresponse. J. Amer. Statist. Assoc. 105, 1265–1275.
Little, R.J.A., Rubin, D.B., 2002. Statistical Inference with Missing Data, second ed.. In: Wiley Series in Probability and Statistics.
Morikawa, K., Kim, J.K., 2016. Semiparametric adaptive estimation with nonignorable nonresponse data. arXiv:1612.09207.
Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57, 1027–1057.
Qin, J., Leung, D., Shao, J., 2002. Estimation with survey data under nonignorable nonresponse or informative sampling. J. Amer. Statist. Assoc. 97, 193–200.
Qin, J., Zhang, B., 2007. Empirical-likelihood-based inference in missing response problems and its application in observational studies. J. R. Stat. Soc. Ser. B Stat. Methodol. 69, 101–122.
Robins, J.M., Rotnitzky, A., Zhao, L.P., 1994. Estimation of regression coefficients when some regressors are not always observed. J. Amer. Statist. Assoc. 89, 846–866.
Rotnitzky, A., Robins, J., 1997. Analysis of semi-parametric regression models with non-ignorable non-response. Stat. Med. 16, 81–102.