Harvard University
Harvard University Biostatistics Working Paper Series

Year 2015 *Paper* 197

On Varieties of Doubly Robust Estimators Under Missing Not at Random With an Ancillary Variable

Wang Miao* Eric Tchetgen Tchetgen†

*Beijing University, mwfy@pku.edu.cn
†Harvard University, etchetge@hsph.harvard.edu

Biometrika (2015), *pp.* 1–8

On Varieties of Doubly Robust Estimators Under Missing Not at Random With an Ancillary Variable

BY WANG MIAO
Beijing International Center for Mathematical Research, Beijing University, Beijing 100871, P.R.C.
mwfy@pku.edu.cn

AND ERIC TCHETGEN TCHETGEN
Department of Biostatistics, Harvard University, Boston, Massachusetts 02115, U.S.A.
etchetge@hsph.harvard.edu

SUMMARY

Suppose we are interested in the mean of an outcome variable missing not at random. Suppose however that one has available an *ancillary variable*, which is associated with the outcome but independent of the missingness process conditional on covariates and the possibly unobserved outcome. Such an ancillary variable may be a proxy or a mismeasured version of the outcome available for all individuals. We have previously established necessary and sufficient conditions for identification of the full data law in such setting, and have described various semiparametric estimators including a doubly robust estimator of the outcome mean. Here, we extend these results and we propose two alternative doubly robust estimators for the outcome mean. The estimators can be viewed as extensions of analogous methods under missing at random, but enjoy different properties.

Some key words: Ancillary variable; Doubly robust; Missing not at random.

1. INTRODUCTION

Doubly robust methods are designed to mitigate estimation bias due to model misspecification in observational studies and imperfect experiments. Such methods have grown in popularity in recent years for estimation with missing data and other forms of coarsening problems (Robins et al., 1994; Scharfstein et al., 1999; Van der Laan & Robins, 2003; Bang & Robins, 2005; Tsiatis, 2007). There currently exist various constructions of doubly robust estimators for the mean of an outcome that is missing at random, i.e. the missingness process only depends on the fully observed covariates. For a review, see Kang & Schafer (2007). In contrast, for data missing not at random, difficulty of identification has dogged the heel of all estimation methods, and doubly robust estimation is far more challenging. Although no general identification results are available for data missing not at random, under specific model assumptions, one may identify the joint distribution of the full data. Building on earlier work by DHaultfoeuille (2010), Wang et al. (2014) and Zhao & Shao (2014), Miao et al. (2015) made use of an *ancillary variable* to establish a general identification framework. An ancillary variable is associated with the outcome but independent of the missingness process conditional on covariates and the outcome. Such a variable may be available in many empirical studies, where a fully observed proxy or a mismeasured version of the outcome may be available. For example, in a study of mental health of children in Connecticut (Zahner et al., 1992; Ibrahim et al., 2001), researchers were interested in evaluating the prevalence of students sampled from a metropolitan center with abnormal psychopathological status based on their teacher's assessment, which was subject to missingness. Interestingly, a separate parent report on the psychopathology of the child was available for all children in the study. Such a report is a proxy for the teacher's assessment, however, it is unlikely to be related to the teacher's response rate conditional on covariates and her assessment of the student, in which case, the parental assessment constitutes a valid ancillary variable. Other examples can be found in Miao et al. (2015), Wang et al. (2014) and Zhao & Shao (2014).

Throughout, we let Y denote the outcome, R is its missingness indicator with $R = 1$ if Y is observed, otherwise $R = 0$, and X denotes fully observed covariates. Suppose that one has also fully observed a variable Z that satisfies the following conditions of an ancillary variable

Assumption 1. (i)$Z \not\!\perp\!\!\!\perp Y \mid X$; (ii) $Z \perp\!\!\!\perp R \mid (Y, X)$.

Assumption 1 formalizes the idea that, as a proxy of Y, Z only affects missingness through its association with Y. The observed data are n independent and identically distributed samples of (Z, RY, R, X), with some values of Y missing. We use the symbols $\mathbb{E}$ and $\mathbb{P}$ to denote expectation and empirical mean, respectively.

In their seminal work, DHaultfoeuille (2010), Wang et al. (2014), Zhao & Shao (2014) and Miao et al. (2015) studied identification of several semiparametric and nonparametric models with a valid ancillary variable. Miao et al. (2015) presented a brief review of such identification problem, and gave necessary and sufficient conditions for identification that are convenient to verify in practice. Particularly, if the outcome is binary, the joint distribution of the full data is always identifiable with a valid ancillary variable. For a continuous outcome, identification requires at least one continuous ancillary variable. But even then, additional conditions are needed to guarantee identification. For instance, based on a pattern-mixture parametrization (Little, 1993), consider the following location-scale models for the outcome density given R

$$P(y|x, z, r) = \frac{1}{\sigma_r(z, x)} P_r \left\{ \frac{y - \mu_r(z, x)}{\sigma_r(z, x)} \right\}, \quad r = 0, 1, \tag{1}$$

with unrestricted functions μ_r and σ_r, and density functions P_r. Under certain regularity conditions summarized in the Appendix, Miao et al. (2015) proved identification for such models. The class includes many commonly-used models, for instance, Gaussian models, and thus essentially states that lack of identification is not an issue in many familiar situations.

Regarding estimation, DHaultfoeuille (2010) presented a nonparametric estimation method based on kernel smoothing, which for reasonable performance requires an unrealistic large sample size with moderate to large covariate dimension. Wang et al. (2014) proposed inverse probability weighted estimation, and Zhao & Shao (2014) studied pseudo-likelihood estimation based on generalized linear models. Their methods will be biased if the required model for propensity score or outcome regression is misspecified. In contrast, a doubly robust approach is consistent under partial misspecification. Specifically, Miao et al. (2015) developed doubly robust estimation based on a three-part model for the full data: a baseline outcome model for the joint distribution of the outcome and the ancillary variable in complete-cases; a baseline propensity score model for response probability evaluated at a baseline value of the outcome; and a log odds ratio model encoding the association of the outcome and the missingness mechanism. Provided correct specification of the log odds ratio model, the doubly robust estimator is consistent if either the baseline outcome model or the baseline propensity score model is correctly specified. However, as shown herein, the construction of a doubly robust estimator under the parametrization of Miao et al. (2015) is not unique. In this paper, we develop two alternative doubly robust estimators of the outcome mean. As we show, the estimators enjoy different properties, and we briefly compare them both in theory and via simulations reported in the Supplementary Material.

2. Doubly robust estimators

Following the parametrization of Miao et al. (2015), we factorize the conditional density function of (Z, Y, R) given X as

$$P(z, y, r|x) = c(x)\exp\{(1-r)OR(y|x)\}P(r|y=0,x)P(z,y|r=1,x), \tag{2}$$

where $c(x)$ is the normalizing constant given x; $P(r|y=0,x)$ is referred to as the baseline propensity score model, i.e. the response probability evaluated at $y=0$; $P(z,y|r=1,x)$ is the density function of (Z,Y) conditional on X among complete cases, and is referred to as the baseline outcome density; $OR(y|x)$ is log of the conditional odds ratio function relating Y and R given X defined as below

$$OR(y|x) = \log\frac{P(r=0|y,x)P(r=1|y=0,x)}{P(r=0|y=0,x)P(r=1|y,x)}.$$

The following equations formulate the main parametrization we shall use throughout for estimation, also see Miao et al. (2015).

$$P(r=1|y,x) = \frac{P(r=1|y=0,x)}{P(r=1|y=0,x)+\exp\{OR(y|x)\}\{P(r=0|y=0,x)\}}, \tag{3}$$

$$P(z,y|r=0,x) = \frac{\exp\{OR(y|x)\}P(z,y|r=1,x)}{\mathbb{E}[\exp\{OR(Y|x)\}|r=1,x]}, \tag{4}$$

$$\mathbb{E}(Y|r=0,x) = \frac{\mathbb{E}[\exp\{OR(Y|x)\}Y|r=1,x]}{\mathbb{E}[\exp\{OR(Y|x)\}|r=1,x]}. \tag{5}$$

For estimation, we specify separate parametric models $OR(y|x;\gamma)$, $P(r=1|y=0,x;\alpha)$, $P(z,y|r=1,x;\beta)$. Throughout, we suppose that the log odds ratio model is correctly specified. The above equations are used to evaluate the inverse probability weights and conditional expectations based on the posited parametric models. The propensity score $P(r=1|x,y;\alpha,\gamma)$, and its inverse, i.e. the inverse probability weight function $W(x,y;\alpha,\gamma) = 1/P(r=1|x,y;\alpha,\gamma)$, are determined by the baseline propensity score model $P(r=1|x,y=0;\alpha)$ and the log odds ratio model $OR(y|x;\gamma)$ as in equation (3). The conditional outcome mean among the subset of the population with missing outcome, i.e. $\mathbb{E}(Y|r=0,x;\beta,\gamma)$ is determined by the baseline outcome model and the log odds ratio model as in equation (5).

We first estimate the nuisance parameters (α,β,γ) by solving

$$\mathbb{P}\{S(z,y,x,r;\widehat{\beta})\} = 0, \tag{6}$$

$$\mathbb{P}[\{W(x,y;\widehat{\alpha},\widehat{\gamma})r-1\}\{E_0(z,x;\widehat{\beta},\widehat{\gamma}),H(x)^T\}^T] = 0, \tag{7}$$

where $H(x)$ is a user-specified vector function, for instance $H(x) = (1, x^T)^T$; $E_0(z,x;\widehat{\beta},\widehat{\gamma}) = z - \mathbb{E}(Z|r=0,x;\widehat{\beta},\widehat{\gamma})$, with the conditional expectation evaluated according to the conditional distribution of equation (4); and $S(z,y,x,r;\widehat{\beta})$ is the estimating equation for the baseline outcome model, for instance, the score function $S(z,y,x,r;\beta) = \partial\log\{P(z,y|r=1,x;\beta)\}/\partial\beta$.

Based on above estimators of the nuisance parameters, we construct three different estimators for the outcome mean that are consistent if together with the log odds ratio model, either the baseline outcome model or the baseline propensity score model is correctly specified.

Regression estimator with residual bias correction. Let $M_0(x;\widehat{\beta},\widehat{\gamma}) = \mathbb{E}(Y|r=0,x;\widehat{\beta},\widehat{\gamma})$, we use the weighted residual to correct bias of $M_0(x;\widehat{\beta},\widehat{\gamma})$

$$\widehat{\mu}_{REG-RBC} = \mathbb{P}[W(x,y;\widehat{\alpha},\widehat{\gamma})r\{y - M_0(x;\widehat{\beta},\widehat{\gamma})\} + M_0(x;\widehat{\beta},\widehat{\gamma})].$$

This class of estimators was previously described by Miao et al. (2015).

Horvitz-Thompson estimator with extended weights. The estimator employs an extended baseline propensity score model and an extended weight function. The extended baseline propensity score model with unknown parameter ϕ satisfies $P_{ext}(r=1|y=0,x;\phi) = P(r=1|y=0,x;\widehat{\alpha})$ only at $\phi=0$. For example, we can specify

$$P_{ext}(r=1|y=0,x;\phi) = \frac{P(r=1|y=0,x;\widehat{\alpha})}{P(r=1|y=0,x;\widehat{\alpha}) + \exp\{\phi g(x)\}P(r=0|y=0,x;\widehat{\alpha})},$$

with user-specified scalar function $g(x)$. The extended weight function $W_{ext}(x,y;\phi)$ is similarly defined as in equation (3), with the baseline propensity score model replaced with the extended one specified above. We estimate the nuisance parameter ϕ of the extended model by solving the following equation

$$\mathbb{P}[\{W_{ext}(x,y;\widehat{\phi})r - 1\}\{M_0(x;\widehat{\beta},\widehat{\gamma}) - \widehat{\mu}_{REG}\}] = 0, \tag{8}$$

with $\widehat{\mu}_{REG} = \mathbb{P}\{(1-r)M_0(x;\widehat{\beta},\widehat{\gamma}) + ry\}$. Horvitz-Thompson estimator with extended weights is

$$\widehat{\mu}_{HT-EXT} = \mathbb{P}\left\{\frac{W_{ext}(x,y;\widehat{\phi})r}{\mathbb{P}\{W_{ext}(x,y;\widehat{\phi})r\}}y\right\}.$$

Regression estimator with an extended outcome model. The estimator involves an extended outcome model $M_{0,ext}(x;\psi)$ with parameter ψ satisfying $M_{0,ext}(x;\psi) = M_0(x;\widehat{\beta},\widehat{\gamma})$ at only $\psi=0$. If $M_0(x;\widehat{\beta},\widehat{\gamma}) = \lambda\{Q(x;\widehat{\beta},\widehat{\gamma})\}$ for some inverse link function λ and some function Q, we can specify $M_{0,ext}(x;\psi) = \lambda\{Q(x;\widehat{\beta},\widehat{\gamma}) + \psi q(x)\}$ with user-specified scalar function $q(x)$. We estimate ψ by solving

$$\mathbb{P}[\{W(x,y;\widehat{\alpha},\widehat{\gamma}) - 1\}r\{y - M_{0,ext}(x;\widehat{\psi})\}] = 0, \tag{9}$$

The regression estimator with extended outcome model is

$$\widehat{\mu}_{REG-EXT} = \mathbb{P}\{(1-r)M_{0,ext}(x;\widehat{\psi}) + ry\}.$$

In special situations, the three estimators can be equivalent. For example, if the components of $H(x)$ include a constant function and $M_0(x;\widehat{\beta},\widehat{\gamma})$ simultaneously, then $\widehat{\mu}_{REG-RBC}$ equals $\widehat{\mu}_{HT-EXT}$ exactly. But in general, they will differ in characteristics, such as bias under misspecification of both baseline models, which we further consider in the next section. According to the next theorem, these three estimators are in fact doubly robust.

THEOREM 1. *Suppose that the log odds ratio model $OR(y|x;\gamma)$ is correctly specified, and that the probability limit of estimating equations* (6)*,* (7)*,* (8) *and* (9) *has a unique solution, then the estimators $\widehat{\mu}_{REG-RBC}$, $\widehat{\mu}_{HT-EXT}$ and $\widehat{\mu}_{REG-EXT}$ are consistent if either $P(z,y|r=1,x;\beta)$ or $P(r=1|y=0,x;\alpha)$ is correctly specified.*

The extended estimators not only provide double robustness, but also provide us with a straightforward strategy to check if the working models are correct. We prove in the Appendix that if the baseline propensity score model is correct, $\widehat{\phi}$ converges to 0, and if the baseline outcome model is correct, $\widehat{\psi}$ converges to 0 in probability. Therefore, one may use this property to assess whether the working models are correctly specified by checking whether these parameters are within sampling variability from zero. The power of the proposed goodness-of-fit test is explored in a simulation study given in the Supplementary Material.

Note that all three doubly robust estimators rely on a correct log odds ratio model. As discussed in Miao et al. (2015), we view this assumption as essentially necessary. Since inference about the law of Y is only possible to the extent that one can untangle the selection process from the full data law, to do so requires an accurate evaluation of their dependence, which is captured by the log odds ratio function $OR(y|x;\gamma)$. To the best of our knowledge, with the exception of Miao et al. (2015), previous doubly robust estimators for missing data have assumed that this log odds ratio is known exactly, either to be equal to the null value of 0 under missing at random (Bang & Robins, 2005; Tsiatis, 2007; Van der Laan & Robins, 2003), or to be of a known functional form with no unknown parameters as in Vansteelandt et al. (2007) and Robins et al. (2008). Therefore, we have in fact relaxed these more stringent assumptions.

3. Relation to previous doubly robust estimators and comparisons

Previous doubly robust estimators under missing at random can be viewed as special cases of our estimators. Under missing at random, $OR(y|x) = 0$, $P(r = 1|x, y = 0) = P(r = 1|x)$, and $\mathbb{E}(Y|x, r = 0) = \mathbb{E}(Y|x)$. Thus, the inverse probability weight function $W(x;\alpha) = 1/P(r = 1|x;\alpha)$ and the outcome model $M(x;\beta) = \mathbb{E}(Y|x;\beta) = M_0(x;\beta) = \mathbb{E}(Y|r = 0, x;\beta)$ do not depend on the log odds ratio. The estimator $\widehat{\mu}_{REG-RBC}$ corresponds to the regression estimator with residual bias correction of Kang & Schafer (2007): $\widehat{\mu}^{MAR}_{REG-RBC} = \mathbb{P}[W(x;\widehat{\alpha})r\{y - M(x;\widehat{\beta})\}] + \mathbb{P}\{M(x;\widehat{\beta})\}$. The estimator $\widehat{\mu}_{HT-EXT}$ corresponds to Horvitz-Thompson doubly robust estimator $\widehat{\mu}^{MAR}_{HT-EXT} = \mathbb{P}[W_{ext}(x;\widehat{\phi})r/\mathbb{P}\{W_{ext}(x;\widehat{\phi})r\}y]$, proposed by Robins et al. (2007). They implemented an extended logistic propensity score model *logit* $P_{ext}(r = 1|x;\phi) = (1, x^T)^T\widehat{\alpha} + \phi g(x)$ with previously obtained $\widehat{\alpha}$ and a user-specified scalar function $g(x)$; and they estimated ϕ by solving $\mathbb{P}[\{W_{ext}(x;\widehat{\phi})r - 1\}\{M(x,\widehat{\beta}) - \mathbb{P}[M(x;\widehat{\beta})]\}] = 0$. The estimator $\widehat{\mu}_{REG-EXT}$ corresponds to the regression doubly robust estimator $\widehat{\mu}^{MAR}_{REG-EXT} = \mathbb{P}\{M_{ext}(x;\widehat{\psi})\}$ proposed by Robins et al. (2007), with the extended outcome model $M_{ext}(x;\widehat{\psi})$ satisfying $\mathbb{P}[W(x;\widehat{\alpha})r\{y - M_{ext}(x;\widehat{\psi})\}] = 0$, and implicitly $\mathbb{P}[r\{y - M_{ext}(x;\widehat{\psi})\}] = 0$ under missing at random.

The three doubly robust estimators for data missing not at random enjoy some of the properties of their analogous versions under missing at random. The estimator $\widehat{\mu}_{HT-EXT}$ is a convex combination of the observed outcome values. It satisfies the property of boundedness (Robins et al., 2007; Tan, 2010) that the estimator falls in the parameter space for the outcome mean almost surely. Such estimators are preferred when the inverse probability weights are highly variable, because they rule out estimates outside the sample space. Boundedness is not guaranteed for $\widehat{\mu}_{REG-RBC}$. If the range of $M_{0,ext}(x;\psi)$ is contained in the sample space of the outcome, $\widehat{\mu}_{REG-EXT}$ also satisfies the boundedness condition, however, this does not hold in general. For example, if the outcome is continuous, and $M_{0,ext}(x;\psi) = M_0(x;\widehat{\beta},\widehat{\gamma}) + \psi$, the range of $\widehat{\mu}_{REG-EXT}$ may be beyond the sample space of the outcome.

Our approach requires an ancillary variable to identify the log odds ratio model. But it is not necessary for data missing at random because the joint distribution of the full data is always identified. Apart from that, the proposed estimators have improvement on bias when both models are misspecified, due to differences in estimation of nuisance parameters. The bias of $\widehat{\mu}_{REG-RBC}$ can be written as

$$Bias_{REG-RBC} = \mathbb{P}[\{W(x, y; \widehat{\alpha}, \widehat{\gamma})r - 1\}\{y - M_0(x; \widehat{\beta}, \widehat{\gamma})\}],$$

and that of $\widehat{\mu}_{REG-EXT}$ has the same form with $M_0(x; \widehat{\beta}, \widehat{\gamma})$ replaced with $M_{0,ext}(x; \widehat{\psi})$. It is driven by the degree of misspecification of both $W(x, y; \widehat{\alpha}, \widehat{\gamma})$ and $M_0(x; \widehat{\beta}, \widehat{\gamma})$. As pointed out by Robins et al. (2007) and Vermeulen & Vansteelandt (2014), without further restrictions on the inverse probability weights, the above bias gets inflated in regions with small propensity scores or large weights. However, as the components of $H(x)$ in equation (7) include a constant function, it implies that $\mathbb{P}\{W(x, y; \widehat{\alpha}, \widehat{\gamma})r\} = 1$. Such condition in fact restricts variability of the inverse probability weights to the extent of bounded expectation. Thus, the bias of $\widehat{\mu}_{REG-RBC}$ and $\widehat{\mu}_{REG-EXT}$ does not explode with large weights.

We briefly compare the relative performance of the three doubly robust estimators via a simulation study (see simulation results in the Supplement Material). From the simulation study, the three doubly robust estimators are consistent if either of the baseline models is correct, but they are biased if neither baseline model is correct. If the baseline outcome model is correct, the parameter of the extended outcome model, $\widehat{\psi}$ is close to 0; and if the baseline propensity score model is correct, the parameter of the extended weight model, $\widehat{\phi}$ is close to 0. Based on asymptotic normality theory for estimation equations, we also test the nulls $\mathbb{H}_0: \ \phi = 0$ and $\mathbb{H}_0: \ \psi = 0$ under level 0.05 and compute their empirical type I error and power. The results show type I error approximating 0.05 if the required baseline propensity score model or baseline outcome model is correct, that is, ϕ or ψ is supposed to be 0. In the settings of the simulation, if the required model is incorrect, such tests have great power close to one in moderate to large samples. As a conclusion, we recommend that one conducts the proposed hypothesis tests to check for severe model misspecification of the baseline models in practice.

4. DISCUSSION

We briefly note that the doubly robust methods described in this work can be further extended to situations focusing on other functionals than the outcome mean, such as a parameter δ solving a full data estimating equation $\mathbb{E}\{U(z, y, x, r; \delta)\} = 0$. This is essentially achieved by replacing Y with U wherever it occurs in the proposed estimating equations and solving the doubly robust estimating equation for the parameter of interest upon setting its mean to zero. The methods also have potential application in other related topics, for example, longitudinal data analysis and causal inference. Their use in such settings is the topic of future study.

ACKNOWLEDGEMENT

The work is partially supported by the China Scholarship Council, and funding from the National Institute of Health.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes the proof of a lemma and simulation studies.

Appendix

Proof of Theorem 1

In order to prove Theorem 1, we need the following lemma which we prove in the Supplementary Material.

Lemma 1. *Suppose that the log odds ratio model is correct, and that the probability limit of equations* (6) *and* (7) *has a unique solution. For any square integrable vector function* $D(z, y, x)$*, scalar function* $V(x)$*, and* $(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})$ *solving equations* (6) *and* (7)*,*

(1) if $P(r = 1|y = 0, x; \alpha)$ *is correct, then* $\mathbb{P}[\{W(x, y; \widehat{\alpha}, \widehat{\gamma})r - 1\}D(z, y, x)]$ *converges to* 0*;*
(2) if $P(z, y|r = 1, x; \beta)$ *is correct, then* $\mathbb{P}[r \exp\{OR(y|x; \widehat{\gamma})\}V(x)\{D(z, y, x) - \mathbb{E}[D(z, y, x)|r = 0, x; \widehat{\beta}, \widehat{\gamma}]\}]$ *converges to* 0*;*
(3) if either of the baseline models is correct, then $\mathbb{P}[\{W(x, y; \widehat{\alpha}, \widehat{\gamma})r - 1\}\{D(z, y, x) - \mathbb{E}[D(z, y, x)|r = 0, x; \widehat{\beta}, \widehat{\gamma}]\}]$ *converges to* 0*.*

Proof of Theorem 1. Suppose that the log odds ratio model is correctly specified, and that the probability limit of the estimating equations has a unique solution.

1. Double robustness of $\widehat{\mu}_{REG-RBC}$. If either of the baseline models is correct, from (3) of Lemma 1, $\mathbb{P}[\{W(x, y; \widehat{\alpha}, \widehat{\gamma})r - 1\}\{y - \mathbb{E}(y|r = 0, x; \widehat{\beta}, \widehat{\gamma})\}]$ converges to 0. For $M_0(x; \widehat{\beta}, \widehat{\gamma}) = \mathbb{E}(y|r = 0, x; \widehat{\beta}, \widehat{\gamma})$, $\mathbb{P}[W(x, y; \widehat{\alpha}, \widehat{\gamma})r\{y - M_0(x; \widehat{\beta}, \widehat{\gamma})\} + M_0(x; \widehat{\beta}, \widehat{\gamma})]$ must converge to the true outcome mean. Therefore, $\widehat{\mu}_{REG-RBC}$ is doubly robust.
2. Double robustness of $\widehat{\mu}_{HT-EXT}$. From (1) of Lemma 1, if the baseline propensity score model is correct, $\mathbb{P}[\{W_{ext}(x, y; \phi = 0)r - 1\}\{M_0(x; \widehat{\beta}, \widehat{\gamma}) - \widehat{\mu}_{REG}\}] = \mathbb{P}[\{W(x, y; \widehat{\alpha}, \widehat{\gamma})r - 1\}\{M_0(x; \widehat{\beta}, \widehat{\gamma}) - \widehat{\mu}_{REG}\}]$ converges to 0, i.e. $\phi = 0$ is a solution of the probability limit of equation (8). Thus, the solution of equation (8), $\widehat{\phi}$ converges to 0, and $\lim_{n\to+\infty} \mathbb{P}\{W_{ext}(x, y; \widehat{\phi})r\} = 1$, $\lim_{n\to+\infty} \mathbb{P}\{W_{ext}(x, y; \widehat{\phi})ry\} = \lim_{n\to+\infty} \mathbb{P}\{W(x, y; \widehat{\alpha}, \widehat{\gamma})ry\} = \mathbb{E}(Y)$. If the baseline outcome model is correct, $\mathbb{P}[(1 - r)\{y - M_0(x; \widehat{\beta}, \widehat{\gamma})\}]$ converges to 0; $\widehat{\mu}_{REG} = \mathbb{P}[(1 - r)M_0(x; \widehat{\beta}, \widehat{\gamma}) + ry]$ converges to the true outcome mean; and $\mathbb{P}(y - \widehat{\mu}_{REG})$ converges to 0. By definition of the extended weight function, $\{W_{ext}(x, y; \widehat{\phi}) - 1\}r = r \exp\{OR(y|x; \widehat{\gamma})\}V(x)$ with $V(x) = P_{ext}(r = 0|y = 0, x; \widehat{\phi})/P_{ext}(r = 1|y = 0, x; \widehat{\phi})$. From (2) of Lemma 1, $\mathbb{P}[\{W_{ext}(x, y; \widehat{\phi}) - 1\}r\{y - M_0(x; \widehat{\beta}, \widehat{\gamma})\}]$ converges to 0, and thus, $\mathbb{P}[\{W_{ext}(x, y; \widehat{\phi})r - 1\}\{y - M_0(x; \widehat{\beta}, \widehat{\gamma})\}]$ converges to 0. Therefore,
$$\begin{aligned}\widehat{\mu}_{HT-EXT} = {} & 1/\mathbb{P}\{W_{ext}(x, y; \widehat{\phi})r\} \cdot \mathbb{P}[\{W_{ext}(x, y; \widehat{\phi})r - 1\}\{y - M_0(x; \widehat{\beta}, \widehat{\gamma})\}] \\ & +1/\mathbb{P}\{W_{ext}(x, y; \widehat{\phi})r\} \cdot \mathbb{P}[\{W_{ext}(x, y; \widehat{\phi})r - 1\}\{M_0(x; \widehat{\beta}, \widehat{\gamma}) - \widehat{\mu}_{REG}\}] \\ & +1/\mathbb{P}\{W_{ext}(x, y; \widehat{\phi})r\} \cdot \mathbb{P}(y - \widehat{\mu}_{REG}) + \widehat{\mu}_{REG} \qquad \square\end{aligned}$$
converges to the true outcome mean. In a word, $\widehat{\mu}_{HT-EXT}$ is doubly robust.
3. Double robustness of $\widehat{\mu}_{REG-EXT}$. If $P(r = 1|x, y = 0; \alpha)$ is correct, from (1) of Lemma 1, $\mathbb{P}[\{W(x, y; \widehat{\alpha}, \widehat{\gamma})r - 1\}\{y - M_{0,ext}(x; \widehat{\psi})\}]$ converges to 0. Note equation (9), we have $\mathbb{P}[(1 - r)\{y - M_{0,ext}(x; \widehat{\psi})\}]$ converges to 0. Thus, $\widehat{\mu}_{REG-EXT} = \mathbb{P}\{(1 - r)M_{0,ext}(x; \widehat{\psi}) + ry\}$ converges to the true outcome mean. If $P(z, y|r = 1, x; \beta)$ is correct, then $\mathbb{P}[(1 - r)\{y - M_0(x; \widehat{\beta}, \widehat{\gamma})\}]$ converges to 0. Since $\{W(x, y; \widehat{\alpha}, \widehat{\gamma}) - 1\}r = r \exp\{OR(y|x; \widehat{\gamma})\}V(x)$ with $V(x) = P(r = 0|y = 0, x; \widehat{\alpha})/P(r = 1|y = 0, x; \widehat{\alpha})$, from (2) of Lemma 1, $\mathbb{P}[\{W(x, y; \widehat{\alpha}, \widehat{\gamma}) - 1\}r\{y - M_{0,ext}(x; \psi = 0)\}] = \mathbb{P}[\{W(x, y; \widehat{\alpha}, \widehat{\gamma}) - 1\}r\{y - M_0(x; \widehat{\beta}, \widehat{\gamma})\}]$ converges to 0. That is, $\psi = 0$ is a solution of the probability limit of equation (9). Thus, the solution of equation (9), $\widehat{\psi}$ converges to 0, and $\lim_{n\to+\infty} \mathbb{P}\{(1 - r)M_{0,ext}(x; \widehat{\psi}) + ry\} = \lim_{n\to+\infty} \mathbb{P}\{(1 - r)M_0(x; \widehat{\beta}, \widehat{\gamma}) + ry\} = \mathbb{E}(Y)$. Therefore, $\widehat{\mu}_{REG-EXT}$ is doubly robust.

Regularity conditions for identification of model (1)

According to Theorem 5 of Miao et al. (2015), with a continuous outcome and a continuous ancillary variable, model (1) is identifiable if it satisfies the following regularity conditions

(1) the characteristic functions $\phi(t)$ of the density function $P_{r=1}(v)$ satisfies $0 < |\phi(t)| < C\exp(-\delta|t|)$ for $t \in \mathbb{R}$ and some constants $C, \delta > 0$;
(2) $\mu(z,x)$, $\sigma(z,x)$ conditional on x, and $P_{r=1}(v)$ are continuously differentiable; and $\int_{-\infty}^{+\infty} |v \cdot \partial P_{r=1}(v)/\partial v|^2 dv$ is finite;
(3) for some linear and one-to-one mapping $M : P_{r=1}\{(\varepsilon - a)/b\} \longmapsto G(t, a, b)$, and any $b, b' > 0$, $(a, b) \neq (a', b')$, $\lim_{t \to t_0} G(t, a, b)/G(t, a', b') = 0$ or ∞ for some t_0.

Many commonly-used models satisfy the conditions, for instance, the standard normal density function with M being the identity mapping and G being the density function itself, and the standard Student-t density with M being the inverse Fourier transform and G being its characteristic function.

REFERENCES

BANG, H. & ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 962–973.

DHAULTFOEUILLE, X. (2010). A new instrumental method for dealing with endogenous selection. *Journal of Econometrics* **154**, 1–15.

IBRAHIM, J. G., LIPSITZ, S. R. & HORTON, N. (2001). Using auxiliary data for parameter estimation with non-ignorably missing outcomes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **50**, 361–373.

KANG, J. D. & SCHAFER, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical science* , 523–539.

LITTLE, R. J. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125–134.

MIAO, W., TCHETGEN TCHETGEN, E. & GENG, Z. (2015). Identification and Doubly Robust Estimation of Data Missing Not at Random With an Ancillary Variable. *ArXiv:1509.02556* .

ROBINS, J., LI, L., TCHETGEN TCHETGEN, E., VAN DER VAART, A. et al. (2008). Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and Statistics: Essays in Honor of David A. Freedman*. Institute of Mathematical Statistics, pp. 335–421.

ROBINS, J., SUED, M., LEI-GOMEZ, Q. & ROTNITZKY, A. (2007). Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable. *Statistical Science* , 544–559.

ROBINS, J. M., ROTNITZKY, A. & ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–866.

SCHARFSTEIN, D. O., ROTNITZKY, A. & ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semi-parametric nonresponse models. *Journal of the American Statistical Association* **94**, 1096–1120.

TAN, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika* **97**, 661–682.

TSIATIS, A. (2007). *Semiparametric Theory and Missing Data*. Springer Science & Business Media.

VAN DER LAAN, M. J. & ROBINS, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer Science & Business Media.

VANSTEELANDT, S., ROTNITZKY, A. & ROBINS, J. (2007). Estimation of regression models for the mean of repeated outcomes under nonignorable nonmonotone nonresponse. *Biometrika* **94**, 841–860.

VERMEULEN, K. & VANSTEELANDT, S. (2014). Biased-reduced doubly robust estimation. *Journal of the American Statistical Association* **accepted**.

WANG, S., SHAO, J. & KIM, J. K. (2014). An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statistica Sinica* **24**, 1097–1116.

ZAHNER, G. E., PAWELKIEWICZ, W., DEFRANCESCO, J. J. & ADNOPOZ, J. (1992). Children's mental health service needs and utilization patterns in an urban community: an epidemiological assessment. *Journal of the American Academy of Child & Adolescent Psychiatry* **31**, 951–960.

ZHAO, J. & SHAO, J. (2014). Semiparametric pseudo likelihoods in generalized linear models with nonignorable missing data. *Journal of the American Statistical Association* **accepted**.

Biometrika (2015), *pp.* 1–6

Printed in Great Britain

Online Supplement to "On Varieties of Doubly Robust Estimators Under Missing Not at Random With an Ancillary Variable"

BY WANG MIAO

Beijing International Center for Mathematical Research, Beijing University, Beijing 100871, China.

mwfy@pku.edu.cn

AND ERIC TCHETGEN TCHETGEN

Department of Biostatistics, Harvard University, Boston, Massachusetts 02115, U.S.A.

etchetge@hsph.harvard.edu

1. PROOF OF LEMMA 1

LEMMA 1. *Suppose that the log odds ratio model is correct, and that the probability limit of equations (6) and (7) has a unique solution. For any square integrable vector function $D(z, y, x)$, scalar function $V(x)$, and $(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})$ solving equations (6) and (7),*

(1) if $P(r = 1|y = 0, x; \alpha)$ is correct, then $\mathbb{P}[\{W(x, y; \widehat{\alpha}, \widehat{\gamma})r - 1\}D(z, y, x)]$ converges to 0;
(2) if $P(z, y|r = 1, x; \beta)$ is correct, then $\mathbb{P}[r \exp\{OR(y|x; \widehat{\gamma})\}V(x)\{D(z, y, x) - \mathbb{E}[D(z, y, x)|r = 0, x; \widehat{\beta}, \widehat{\gamma}]\}]$ converges to 0;
(3) if either of the baseline models is correct, then $\mathbb{P}[\{W(x, y; \widehat{\alpha}, \widehat{\gamma})r - 1\}\{D(z, y, x) - \mathbb{E}[D(z, y, x)|r = 0, x; \widehat{\beta}, \widehat{\gamma}]\}]$ converges to 0.

Proof of Lemma 1. We use starred characters $(\alpha^*, \beta^*, \gamma^*)$ to denote probability limit of the solutions of equations (6) and (7), and $(\alpha^0, \beta^0, \gamma^0)$ to denote true values of the nuisance parameters. Suppose that the log odds ratio model is correct.

(1) If $P(r = 1|x, y; \alpha)$ is correct, then $\mathbb{E}\{W(x, y; \alpha^0, \gamma^0)r - 1|x, y\} = 0$. By the ancillary variable assumption $Z \perp\!\!\!\perp R|(X, Y)$, for any square integrable function $D(z, y, x)$, $\mathbb{E}[\{W(x, y; \alpha^0, \gamma^0)r - 1\}D(z, y, x)|x, y] = 0$, and thus $\mathbb{E}[\{W(x, y; \alpha^0, \gamma^0)r - 1\}D(z, y, x)] = 0$. Therefore, $(\alpha^0, \beta^*, \gamma^0)$ is a solution of the probability limit of equations (6) and (7) with $D(z, y, x) = \{E_0(z, x; \beta^*, \gamma^0), H(x)^T\}^T$. Since the probability limit of equations (6) and (7) have a unique solution, $(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})$ must converge to $(\alpha^0, \beta^*, \gamma^0)$. Thus, for any square integrable function $D(z, y, x)$, $\mathbb{P}[\{W(x, y; \widehat{\alpha}, \widehat{\gamma})r - 1\}D(z, y, x)]$ converges to $\mathbb{E}[\{W(x, y; \alpha^0, \gamma^0)r - 1\}D(z, y, x)]$, which equals 0.
(2) For any square integrable function $D(z, y, x)$, from equation (4) we can prove that

$$\mathbb{E}\{D(z, y, x)|r = 0, x\} = \int\int D(z, y, x)P(z, y|r = 0, x)dzdy$$
$$= \frac{\mathbb{E}[r \exp\{OR(y|x)\}D(z, y, x)|x]}{\mathbb{E}[r \exp\{OR(y|x)\}|x]}.$$

Thus,

$$\mathbb{E}[r\exp\{OR(y|x)\}D(z,y,x)|x] = \mathbb{E}[r\exp\{OR(y|x)\}|x]\cdot\mathbb{E}\{D(z,y,x)|r=0,x\},$$

$$\mathbb{E}[r\exp\{OR(y|x)\}\{D(z,y,x)-\mathbb{E}[D(z,y,x)|r=0,x]\}|x]=0,$$

By the property of pulling out known factors of conditional expectations, for any square integrable scalar function $V(x)$

$$\mathbb{E}[r\exp\{OR(y|x)\}V(x)\{D(z,y,x)-\mathbb{E}[D(z,y,x)|r=0,x]\}|x]=0,$$

and by the law of iterated expectation,

$$\mathbb{E}[r\exp\{OR(y|x)\}V(x)\{D(z,y,x)-\mathbb{E}[D(z,y,x)|r=0,x]\}]=0.$$

If $P(z,y|r=1,x;\beta)$ is correct, then

$$\mathbb{E}[r\exp\{OR(y|x;\gamma^0)\}V(x)\{D(z,y,x)-\mathbb{E}[D(z,y,x)|r=0,x;\beta^0,\gamma^0]\}]=0,$$

$$\mathbb{E}[(1-r)\{D(z,y,x)-\mathbb{E}[D(z,y,x)|r=0,x;\beta^0,\gamma^0]\}]=0.$$

From equation (3), $\{W(x,y;\alpha^*,\gamma^0)-1\}r = r\exp\{OR(y|x;\gamma^0)\}V(x;\alpha^*)$ with $V(x;\alpha^*) = P(r=0|y=0,x;\alpha^*)/P(r=1|y=0,x;\alpha^*)$. Thus,

$$\mathbb{E}[\{W(x,y;\alpha^*,\gamma^0)r-1\}\{D(z,y,x)-\mathbb{E}[D(z,y,x)|r=0,x;\beta^0,\gamma^0]\}]=0,$$

Therefore, if $P(z,y|r=1,x;\beta)$ is correct, $(\alpha^*,\beta^0,\gamma^0)$ is a solution of the probability limit of equations (6) and (7) with $D(z,y,x)=\{E_0(z,x;\beta^0,\gamma^0),H(x)^T\}^T$. Since the probability limit of equations (6) and (7) have a unique solution, $(\widehat{\alpha},\widehat{\beta},\widehat{\gamma})$ converges to $(\alpha^*,\beta^0,\gamma^0)$. Thus, for any square integrable functions $D(z,y,x)$ and $V(x)$,

$$\begin{aligned}&\lim_{n\to+\infty}\mathbb{P}[r\exp\{OR(y|x;\widehat{\gamma})\}V(x)\{D(z,y,x)-\mathbb{E}[D(z,y,x)|r=0,x;\widehat{\beta},\widehat{\gamma}]\}]\\ =\ &\mathbb{E}[r\exp\{OR(y|x;\gamma^0)\}V(x)\{D(z,y,x)-\mathbb{E}[D(z,y,x)|r=0,x;\beta^0,\gamma^0]\}]=0.\end{aligned}$$

(3) If $P(r=1|x,y=0;\alpha)$ is correct, the result is implied by (1). If $P(z,y|r=1,x;\beta)$ is correct, from the proof of (2), $(\widehat{\alpha},\widehat{\beta},\widehat{\gamma})$ converges to $(\alpha^*,\beta^0,\gamma^0)$, and

$$\begin{aligned}&\lim_{n\to+\infty}\mathbb{P}[\{W(x,y;\widehat{\alpha},\widehat{\gamma})r-1\}\{D(z,y,x)-\mathbb{E}[D(z,y,x)|r=0,x;\widehat{\beta},\widehat{\gamma}]\}]\\ =\ &\mathbb{E}[\{W(x,y;\alpha^*,\gamma^0)r-1\}\{D(z,y,x)-\mathbb{E}[D(z,y,x)|r=0,x;\beta^0,\gamma^0]\}]=0, \qquad \square\end{aligned}$$

2. SIMULATION RESULTS FOR A NORMAL OUTCOME

Except for the three doubly robust estimators, in the simulation study, we also consider non-doubly robust estimators presented by Miao et al. (2015) and Wang et al. (2014): an outcome regression based estimator

$$\widehat{\mu}_{reg} = \mathbb{P}\{(1-r)M_0(x;\widehat{\beta},\widehat{\gamma}_{reg}) + ry\},$$

with $(\widehat{\beta},\widehat{\gamma}_{reg})$ solving

$$\begin{aligned}&\mathbb{P}\{S(z,y,x,r;\widehat{\beta})\}=0,\\ &\mathbb{P}[(1-r)\{z-\mathbb{E}(Z|r=0,x;\widehat{\beta},\widehat{\gamma}_{reg})\}]=0;\end{aligned}$$

and an inverse probability weighted estimator

$$\widehat{\mu}_{ipw} = \mathbb{P}\{W(x, y; \widehat{\alpha}_{ipw}, \widehat{\gamma}_{ipw}) ry\},$$

with $(\widehat{\alpha}_{ipw}, \widehat{\gamma}_{ipw})$ solving

$$\mathbb{P}[\{W(x, y; \widehat{\alpha}_{ipw}, \widehat{\gamma}_{ipw}) r - 1\}(z, 1, x^T)^T] = 0.$$

Provided the log odds ratio model is correctly specified, the outcome regression based estimator is consistent only if the baseline outcome model is correct, and the inverse probability weighted estimator is consistent only if the baseline propensity score model is correct (Miao et al., 2015).

We generate a covariate $X \sim N(0, 1)$, and then generate data from conditional density (2) in Section 2 with a linear log odds ratio model:

$$OR(y|x) = -\gamma_0 y - \gamma_1 \frac{y}{0.3 + 0.5x^2},$$

a bivariate normal baseline outcome model and a logistic baseline propensity score model. We consider two choices of the baseline outcome model for data generating:

$$Y|r = 1, x \sim N\{\beta_{10} + \beta_{11}(0.5x + 0.333x^2), 1\}, \quad Z|y, x \sim N\{\beta_{20} + \beta_{21}(x + 2x^2) + \beta_{22}y, 1\},$$

and

$$Y|r = 1, x \sim N(\beta_{10} + \beta_{11}x, 1), \quad Z|y, x \sim N(\beta_{20} + \beta_{21}x^2 + \beta_{22}y, 1).$$

We also consider two choices of the baseline propensity score model:

$$\mathit{logit}\, P(r = 1|y = 0, x) = \alpha_0 + \alpha_1(x - 0.5x^2),$$

and

$$\mathit{logit}\, P(r = 1|y = 0, x) = \alpha_0 + \alpha_1 x.$$

So we have four data generating mechanisms. The parameters are set equal to: $(\alpha_0, \alpha_1) = (0.5, 1)$, $(\beta_{10}, \beta_{11}) = (0.5, 1)$, $(\beta_{20}, \beta_{21}, \beta_{22}) = (-0.5, 2, 3)$, and $(\gamma_0, \gamma_1) = (-0.8, 0.5)$. For these settings, the missing data proportions are between 40% and 50%.

According to Section 1, The models are identifiable as $\beta_{22} \neq 0$. For estimation, we assume the correct log odds ratio model, and always use the working models below to estimate the nuisance parameters:

$$Y|r = 1, x \sim N(\beta_{10} + \beta_{11}x, \sigma_y^2), \quad Z|y, x \sim N(\beta_{20} + \beta_{21}x^2 + \beta_{22}y, \sigma_z^2),$$

with

$$\mathit{logit}\, P(r = 1|y = 0, x) = \alpha_0 + \alpha_1 x.$$

For Horvitz-Thompson estimator with extended weights, we use the following extended baseline propensity score model:

$$\mathit{logit}\, P_{ext}(r = 1|y = 0, x; \phi) = \widehat{\alpha}_0 + \widehat{\alpha}_1 x + \phi x;$$

for regression estimator with an extended outcome model, we use the following extended outcome model:

$$M_{0,ext}(x; \psi) = M_0(x; \widehat{\beta}_{10}, \widehat{\beta}_{11}, \widehat{\gamma}_0, \widehat{\gamma}_1) + \psi x^2,$$

where $M_0(x; \widehat{\beta}_{10}, \widehat{\beta}_{11}, \widehat{\gamma}_0, \widehat{\gamma}_1) = \mathbb{E}(Y|r = 0, x; \widehat{\beta}_{10}, \widehat{\beta}_{11}, \widehat{\gamma}_0, \widehat{\gamma}_1)$, with estimates of nuisance parameters $(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\beta}_{10}, \widehat{\beta}_{11}, \widehat{\gamma}_0, \widehat{\gamma}_1)$ obtained from the above working models.

We simulate 1000 replicates under 500 and 1500 sample sizes for each data generating mechanism and summarize the results in Figures 1, 2, and Table 1. Figure 1 presents the results for estimation of the outcome mean. From (a), (b) and (c) of Figure 1, the three doubly robust estimators are consistent if either the baseline outcome model or the baseline propensity score model is correct. But they are biased if neither model is correct, as shown in (d) of Figure 1. As expected, the outcome regression based estimator is consistent only if the baseline outcome model is correct, and the inverse probability weighted estimator is consistent only if the baseline propensity score model is correct. Figure 1 show robustness of the doubly robust estimator. So we recommend doubly robust estimation to evaluate the mean of a normal outcome.

Figure 2 presents results for parameters of the extended models. If the baseline outcome model is correct, parameter of the extended outcome model, $\widehat{\psi}$ is close to 0, and if the baseline propensity score model is correct, parameter of the extended weight model, $\widehat{\phi}$ is close to 0. We also conduct hypothesis tests to test the nulls $\mathbb{H}_0: \ \phi = 0$, and $\mathbb{H}_0: \ \psi = 0$ under 0.05 level. Such tests are based on asymptotic normality of the estimators $\widehat{\phi}$ and $\widehat{\psi}$, which follows the classical theory for estimating equations that can be found in, for example, Hall (2005). Table 1 presents empirical type I error and power of the hypothesis tests. From the first column, if the baseline propensity score model is correct, i.e. ϕ is supposed to be 0, the empirical type I error of testing the null $\mathbb{H}_0: \ \phi = 0$ approximates the nominal level 0.05 as the sample size increases. From the second column, if the baseline propensity score model is incorrect, the empirical power of testing the null $\mathbb{H}_0: \ \phi = 0$ is closed to one. The results are also true for testing the null $\mathbb{H}_0: \ \psi = 0$ as shown in the third and fourth columns. As a conclusion, we recommend such hypothesis tests to access correctness of the baseline models.

Table 1: Empirical type I error and power

	$\mathbb{H}_0: \ \phi = 0$		$\mathbb{H}_0: \ \psi = 0$	
	Type I error	Power	Type I error	Power
(a)	—	0.844	0.036	—
	—	0.998	0.044	—
(b)	0.022	—	—	0.657
	0.036	—	—	0.946
(c)	0.022	—	0.042	—
	0.031	—	0.058	—
(d)	—	0.644	—	0.345
	—	0.961	—	0.844

Note: Data are generated from four scenarios. The baseline propensity score model is correct in (b) and (c), but incorrect in (a) and (d); the baseline outcome model is correct in (a) and (c), but incorrect in (b) and (d). Columns 1 and 3 present empirical type I error of the tests as the baseline propensity score model or the baseline outcome model is correct, respectively. Columns 2 and 4 present empirical power of the tests as the baseline propensity score model or the baseline outcome model is incorrect, respectively. The symbol "—" indicates inapplicable. The result of each situation includes two rows, of which the first row stands for sample size 500, and the second for 1500.

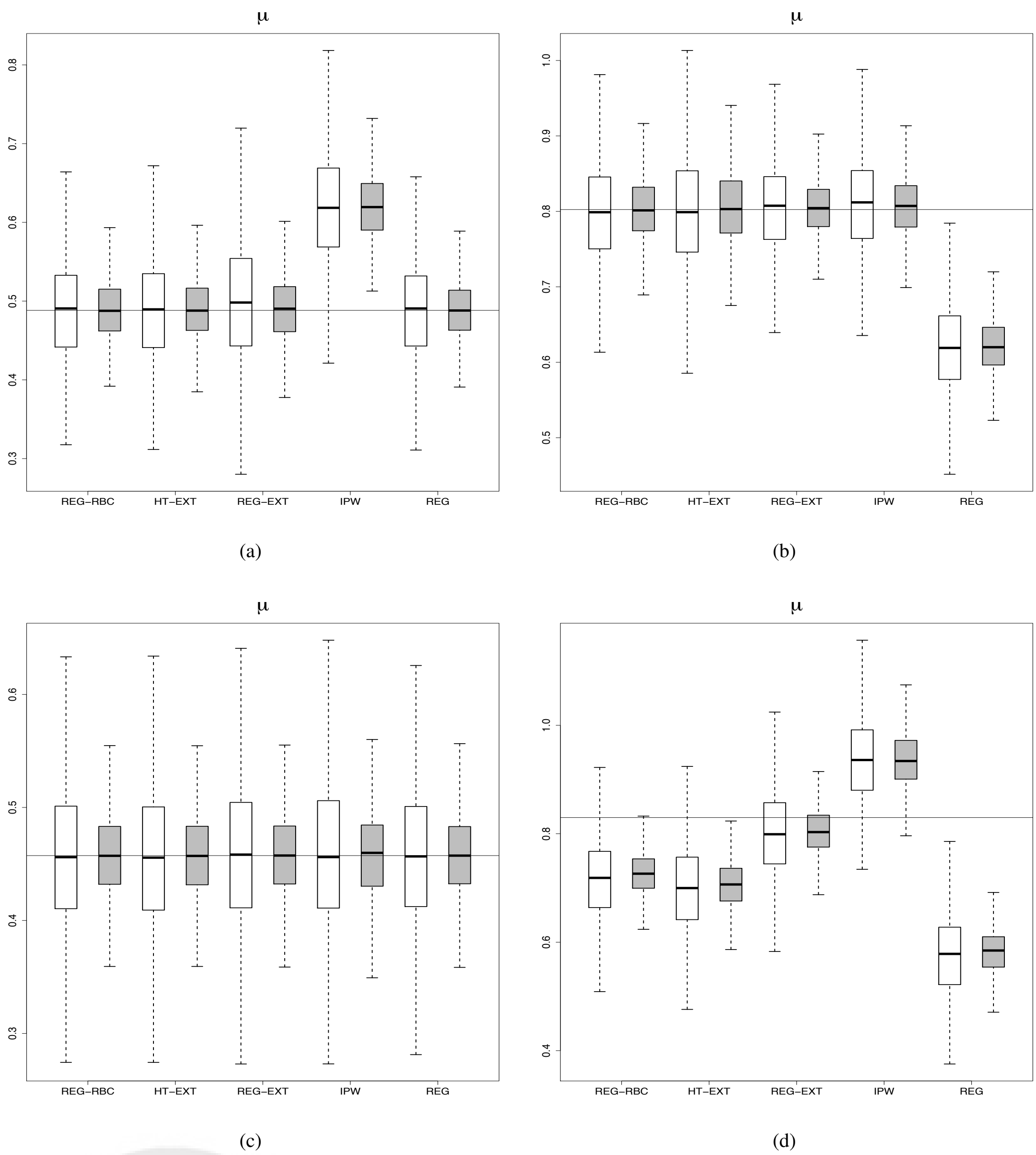

Fig. 1: Boxplots of the estimators for the mean of a normal outcome.

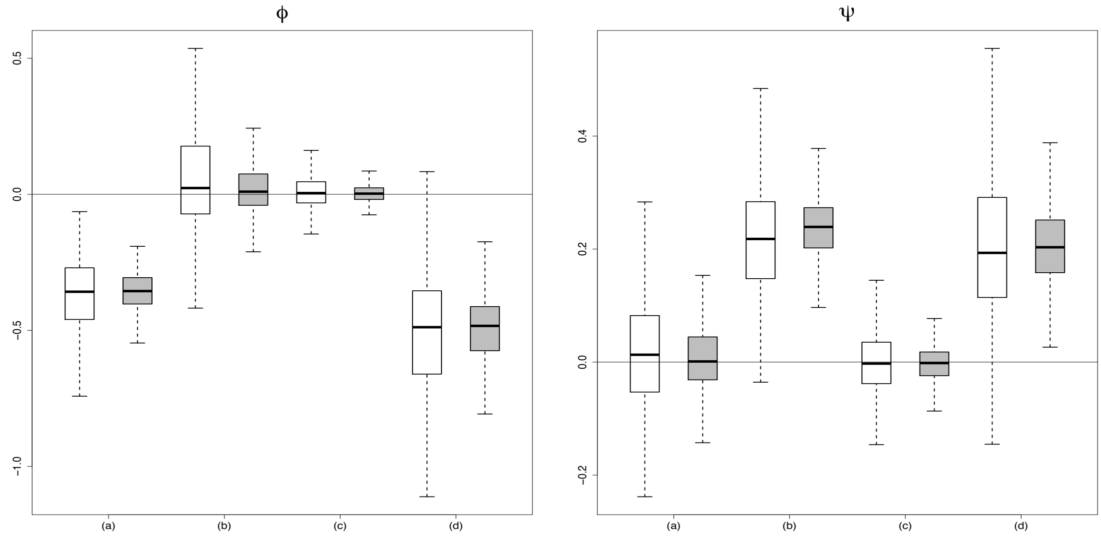

Fig. 2: Boxplots of the estimators for parameters of the extended models.

Note for Figures 1 and 2: Data are generated from four scenarios. The baseline propensity score model is correct in (b) and (c), but incorrect in (a) and (d); the baseline outcome model is correct in (a) and (c), but incorrect in (b) and (d). Data are analyzed with five methods: regression estimation with residual bias correction (REG-RBC), Horvitz-Thompson estimation with extended weights (HT-EXT), regression estimation with an extended outcome model (REG-EXT), inverse probability weighting (IPW), and outcome regression based estimation (REG). In each boxplot, white boxes are for sample size 500, and gray ones for 1500. The horizontal line marks the true value of the outcome mean in Figure 1, and 0 in Figure 2.

REFERENCES

HALL, A. R. (2005). *Generalized Method of Moments*. Oxford University Press Oxford.

MIAO, W., TCHETGEN TCHETGEN, E. & GENG, Z. (2015). Identification and Doubly Robust Estimation of Data Missing Not at Random With an Ancillary Variable. *ArXiv:1509.02556* .

WANG, S., SHAO, J. & KIM, J. K. (2014). An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statistica Sinica* **24**, 1097–1116.