Statistica Sinica Preprint No: SS-2023-0383	
Title	On Doubly Robust Estimation with Nonignorable Missing Data Using Instrumental Variables
Manuscript ID	SS-2023-0383
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0383
Complete List of Authors	Baoluo Sun, Wang Miao and Deshanee S. Wickramarachchi
Corresponding Authors	Baoluo Sun
E-mails	stasb@nus.edu.sg

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

On Doubly Robust Estimation with Nonignorable Missing Data Using Instrumental Variables

Baoluo Sun[1], Wang Miao[2], and Deshanee S. Wickramarachchi[3]

[1]*Department of Statistics and Data Science, National University of Singapore*

[2]*Department of Probability and Statistics, Peking University*

[3]*Department of Statistics, University of Colombo*

Abstract: Suppose we are interested in the mean of an outcome that is subject to nonignorable nonresponse. This paper develops new semiparametric estimation methods with instrumental variables which affect nonresponse, but not the outcome. The proposed estimators remain consistent and asymptotically normal even under partial model misspecifications for two variation independent nuisance components. We evaluate the performance of the proposed estimators via a simulation study, and apply them in adjusting for missing data induced by HIV testing refusal in the evaluation of HIV seroprevalence in Mochudi, Botswana, using interviewer experience as an instrumental variable.

Key words and phrases: Doubly robust estimation, Endogeneous selection, Exclusion restriction, Instrumental variable, Nonignorable nonresponse

1. Introduction

Missing data are ubiquitous in the health and social sciences. Our motivating example concerns a household survey in Mochudi, Botswana to estimate HIV seroprevalence among adults. About 19% of the adults who were contacted for the survey have missing final HIV status, mainly due to refusal to participate in the HIV testing component. The nonresponse is said to be ignorable if, conditional on the fully observed

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

variables, it is independent of the underlying HIV status (Rubin, 1976; Rubin and Little, 2019). In this case, the HIV seroprevalence among respondents is representative of the overall HIV seroprevalence in the population, within strata of the observed variables. Nonetheless, ignorability is a strong assumption which may be untenable in practice; for instance, HIV testing refusal may be entangled with features of the underlying HIV status in the household survey. The problem of nonignorable nonresponse has therefore received much attention in the missing data literature. Robins et al. (2000) described a general class of models which requires *a priori* specification of a selection bias parameter that encodes the residual association of the nonresponse mechanism with the outcome of interest, after adjusting for the fully observed variables. It coincides with the widely adopted exponential tilting model in the special case where this residual association is captured on the log odds ratio scale (Vansteelandt et al., 2007; Kim and Yu, 2011; Miao et al., 2024; Miao and Tchetgen Tchetgen, 2016; Shao and Wang, 2016). When the selection bias or tilting parameter in the exponential tilting model is *a priori* known or can be estimated from external data, functionals of the complete data distribution such as the outcome mean can be estimated based on either the inverse propensity weighting or regression approach (Scharfstein et al., 1999; Vansteelandt et al., 2007; Kim and Yu, 2011). These two approaches can be carefully combined to obtain doubly robust estimators which remain consistent and asymptotically normal if either the propensity score or regression model, but not necessarily both, is correctly specified. Such methods have grown in popularity in recent years for estimation with missing data and other forms of coarsening, as they effectively double one's chances to obtain valid inference (Robins et al., 1994; Scharfstein et al., 1999; Robins and Rotnitzky, 2001; van der Laan and Robins,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

2003; Bang and Robins, 2005; Tsiatis, 2007; Molenberghs et al., 2014; Seaman and Vansteelandt, 2018; Chernozhukov et al., 2022). Furthermore, these methods are also locally semiparametric efficient because they are able to achieve the semiparametric efficiency bound when all the model specifications hold.

Doubly robust methods when the tilting parameter is unknown are far less developed and generally require specific, non-trivial constructions. Miao and Tchetgen Tchetgen (2016) and Miao et al. (2024) developed doubly robust inference by leveraging a shadow variable which affects the outcome but not the nonresponse. Such a variable may be available in many empirical studies, and has played a prominent role in semiparametric estimation with nonignorable nonresponse (Liang and Qin, 2000; Tang et al., 2003; Chen et al., 2009; d'Haultfoeuille, 2010; Wang et al., 2014; Zhao and Shao, 2015; Shao and Wang, 2016; Zhao and Ma, 2022; Li et al., 2023). Another approach involves instrumental variables which affect nonresponse, but not the outcome. The instrumental variable approach has a longstanding tradition in econometrics (Heckman, 1974, 1979; Manski, 1990; Ahn and Powell, 1993; Powell, 1994; Das et al., 2003), and has witnessed renewed interests in the health and social sciences by leveraging certain operational features of a study (Lepkowski et al., 2002; Schräpler, 2004; Nicoletti and Peracchi, 2005; Bärnighausen et al., 2011; Tchetgen Tchetgen and Wirth, 2017; Sun et al., 2018; Marden et al., 2018). In the motivating example, interviewers are randomly deployed prior to the survey, possibly given the values of baseline covariates such as administrative regions. Therefore, interviewer characteristics such as years of experience constitute candidate instruments which likely influence the response rates of individuals contacted for the survey, but are independent of the individuals' underlying outcomes of interest, within strata of

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

the baseline covariates.

Sun et al. (2018) established the semiparametric efficiency theory for estimating the outcome mean with instrumental variables when the tilting parameter is unknown. They also proposed a doubly robust estimator of the outcome mean based on the widely adopted odds ratio factorization of the complete data distribution, in which the outcome density among nonrespondents can be expressed as an exponential tilting of the density among respondents (Robins and Rotnitzky, 2001; Vansteelandt et al., 2007; Kim and Yu, 2011; Miao et al., 2024; Miao and Tchetgen Tchetgen, 2016; Shao and Wang, 2016; Riddles et al., 2016; Malinsky et al., 2022); see Kim and Shao (2021) for a recent review. However, as we show later in the paper, doubly robust inference with instrumental variables is complicated by the fact that, under this widely adopted factorization, the models for different components of the complete data distribution are in fact variationally dependent. Therefore, any choice of one model imposes *a priori* restrictions on the range of possible models for the remaining components, which implies that we may not have two independent opportunities for valid inference. Furthermore, one may in fact rule out the possibility of achieving local semiparametric efficiency due to possible lack of model compatibility, which leaves a gap in the nonignorable missing data literature. The main contribution of our paper is to provide explicit construction of novel doubly robust and locally efficient estimators which resolves these issues through an alternative factorization of the complete data distribution that naturally encodes the instrumental variable assumptions.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

2. Preliminaries

2.1 Model and assumptions

The full data $W = (Y, Z, U)$ has support $\mathcal{W} = (\mathcal{Y} \times \mathcal{Z} \times \mathcal{U})$. Here Y is an outcome of interest, Z an instrumental variable, and $U = (U_1, ..., U_L)$ consists of L measured baseline covariates. Suppose that $X = (Z, U)$ is fully observed, but Y is subject to missingness. Let $R \in \{0, 1\}$ denote the binary random variable indicating missingness status, with $R = 1$ if Y is observed and $R = 0$ otherwise. We assume that the distribution of the complete data (R, W) has density

$$p(r, w) = \pi(w)^r \{1 - \pi(w)\}^{1-r} p(w),$$

with respect to some appropriate dominating measure, where $\pi(w) = p(R = 1 \mid w)$ denotes the extended propensity score which captures the nonresponse mechanism, and $p(w)$ is the full data density. Throughout, we make the following positivity assumption, which is necessary for identification of the complete data distribution and smooth functionals of the latter, and ensures finite asymptotic variance of the proposed estimators (Rotnitzky et al., 1998).

Assumption 1. $\pi(w) > \sigma > 0$ for all $w \in \mathcal{W}$, where σ is a fixed positive constant.

Let $p(r, w; \varphi)$ denote a model for the complete data density indexed by φ, which may be infinite dimensional. We are interested in identifying φ from the observed data distribution, which is captured by $p(x; \varphi)$ and $p(y, R = 1 \mid x; \varphi)$. More formally, the parameter φ is said to be identified from the observed data, if there exists a one-to-one mapping between φ and $\{p(x; \varphi), p(y, R = 1 \mid x; \varphi)\}$. It is well known that φ cannot be identified from the observed data in the absence of further assumptions

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

(Robins and Ritov, 1997; Robins et al., 2000; van der Laan and Robins, 2003). In this paper, we adopt the instrumental variable framework by assuming that Z should affect nonresponse, but not the outcome, within strata of measured baseline covariates.

Assumption 2. $Z \not\perp\!\!\!\perp R \mid U$ (instrumental variable relevance). $Z \perp Y \mid U$ (exclusion restriction).

Assumption 2 is consistent with the identifiability conditions in prior works with instrumental variables (Heckman, 1974, 1979; Manski, 1990; Ahn and Powell, 1993; Powell, 1994; Das et al., 2003; Tchetgen Tchetgen and Wirth, 2017; Sun et al., 2018). The exclusion restriction implies that the model for the full data density can be factorized as $p(w;\psi,\beta,\zeta) = p(y \mid u;\psi)p(z \mid u;\beta)p(u;\zeta)$. Nonetheless, φ cannot be identified from the observed data even with exclusion restriction, as illustrated by the following example.

Example 1. Suppose $\mathcal{Y} = \mathcal{Z} = \{0,1\}$, and there are no baseline covariates. Then we are able to identify the five parameters $p(y, R=1 \mid z)$, $p(Z=1)$ from the observed data, but under assumption 2, there are six unknown parameters $p(Y=1)$, $p(Z=1)$ and $p(R=1 \mid y,z)$, which remains unidentifiable by parameter counting.

Therefore, we need to restrict the candidates for the complete data distribution to an even smaller set. Specifically, a working parametric model is assumed for the selection bias function

$$h(y,x) = \log\left\{\frac{p(R=1 \mid y,x)/p(R=0 \mid y,x)}{p(R=1 \mid Y=0,x)/p(R=0 \mid Y=0,x)}\right\}, \tag{2.1}$$

which encodes our *a priori* belief of the way in which the underlying outcome affects the response mechanism on the log odds ratio scale. This implies the following

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

semiparametric exponential tilting model for the propensity score,

$$\pi(w;\eta,\gamma) = \text{expit}\{\eta(x) + h(y,x;\gamma)\}, \tag{2.2}$$

where $\text{expit}(t) = 1/\{1+\exp(-t)\}$ is the inverse logit function, $\eta(x)$ is an unknown function, and $h(y,x;\gamma)$ is a known function smooth in the finite-dimensional parameter $\gamma \in \mathbb{R}^{p_\gamma}$, which satisfies $h(0,x;\gamma) = 0$ based on the odds ratio representation (2.1). In addition, the parameterization is typically chosen to be such that $h(y,x;0) = 0$, so that $\gamma = 0$ corresponds to the ignorable nonresponse mechanism. A common specification is $h(y,x;\gamma) = \gamma y$ (Kim and Yu, 2011; Shao and Wang, 2016), and in general the selection bias function can be specified more flexibly to allow for dependence of such effects on fully observed covariates.

The missingness is nonignorable if and only if the true parameter value $\gamma_0 \neq 0$. As γ_0 is not identified in the absence of further assumptions, sensitivity analysis has been proposed whereby one conducts inferences assuming $\gamma = \gamma_0$ is completely known and repeats the analysis upon varying the assumed value of γ (Rotnitzky et al., 1998; Scharfstein et al., 1999; Robins et al., 2000; Vansteelandt et al., 2007). Kim and Yu (2011) assumed that γ_0 is known or can be estimated using external data, while Miao et al. (2024) and Shao and Wang (2016) used shadow variables to estimate γ_0. In this paper, we follow the instrumental variable approach of Sun et al. (2018) and assume that the following condition holds for identification.

Assumption 3. The semiparametric exponential tilting model (2.2) is correctly specified. Furthermore, for any given $u \in \mathcal{U}$, the ratio $\pi(w;\eta_1,\gamma_1)/\pi(w;\eta_2,\gamma_2)$ is either a constant or varies with z for any two values (η_1,γ_1) and (η_2,γ_2) of (η,γ).

Note that assumption 3 does not impose further restrictions on the full data

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

density model $p(w;\psi,\beta,\zeta)$. In the supplementary material, we show that assumptions 1–3 are sufficient for identification of the complete data distribution from the observed data.

2.2 Prior works

We briefly review existing methods in the nonignorable missing data literature under semiparametric exponential tilting model (2.2). If $\gamma = \gamma_0$ is known, then the outcome mean $\mu_0 = E(Y)$ can be estimated using either the inverse propensity weighting or regression approach (Scharfstein et al., 1999; Vansteelandt et al., 2007; Kim and Yu, 2011), based on the representations

$$\mu_0 = \mathbb{E}\left\{\frac{RY}{\pi(W)}\right\},$$

or

$$\mu_0 = \mathbb{E}\{RY + (1-R)\mathbb{E}(Y \mid R=0, X)\},$$

respectively. For the inverse propensity weighting approach, the unknown function $\eta(x)$ is nonparametrically identified based on the conditional moment restriction,

$$\mathbb{E}\left\{\frac{R}{\pi(W;\eta,\gamma)} - 1 \middle| X\right\} = 0, \tag{2.3}$$

for each fixed value of γ (Vansteelandt et al., 2007; Kim and Yu, 2011; Shao and Wang, 2016). For the second representation, the odds ratio factorization of the complete data distribution is widely adopted, in which the density among nonrespondents can be expressed as an exponential tilting of the density among respondents,

$$p(y \mid R=0, x;\gamma) = \frac{\exp\{-h(y,x;\gamma)\}p(y \mid R=1,x)}{\int_{\mathcal{Y}} \exp\{-h(t,x;\gamma)\}p(t \mid R=1,x)d\nu(t)}, \tag{2.4}$$

where ν is an appropriate dominating measure (Chen, 2007; Vansteelandt et al., 2007; Tchetgen Tchetgen et al., 2010; Kim and Yu, 2011; Riddles et al., 2016). Thus, if

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

$\gamma = \gamma_0$ is known or can be estimated using external data, then estimation of μ_0 entails estimation of either $\eta(x)$ or $p(y \mid R = 1, x)$. When X is high-dimensional or contains numerous continuous components, we can specify parametric or semiparametric models for these unknown functions. In the absence of further restrictions, the parameterization of the models for $\eta(x)$ and $p(y \mid R = 1, x)$ are variationally independent (Chen, 2007). This provides the basis for doubly robust inference about the outcome mean if either the model for $\eta(x)$ or $p(y \mid R = 1, x)$ is correctly specified (Vansteelandt et al., 2007). In particular, the value $\gamma = 0$ corresponds to doubly robust inference under ignorable nonresponse.

If the true value γ_0 is unknown, Miao et al. (2024), Miao and Tchetgen Tchetgen (2016) and Sun et al. (2018) developed doubly robust inference for μ_0 under a similar odds ratio factorization of the complete data distribution. However, the possibility for genuine doubly robust inference with instrumental variables is complicated by the fact that the models for $\eta(x)$ and $p(y \mid R = 1, x)$ are variationally dependent under the exclusion restriction of assumption 2. To characterize this dependency, the propensity score $\tilde{\pi}(x) = p(R = 1 \mid x)$ can be expressed under model (2.2) as $\tilde{\pi}(x) = 1/\{1 + \iota(x)\}$, where $\iota(x) = \mathbb{E}[\exp\{-\eta(X) - h(Y, X; \gamma)\} \mid R = 1, X = x]$ (Kim and Shao, 2021, Lemma 8.1). The outcome density conditional on fully observed values is given by

$$p(y \mid x) = p(y \mid R = 1, x)\tilde{\pi}(x) + p(y \mid R = 0, x)\{1 - \tilde{\pi}(x)\}. \tag{2.5}$$

Assumption 2 imposes the additional exclusion restriction that $p(y \mid z, u) = p(y \mid u)$ for all $w \in \mathcal{W}$, which induces variation dependency between the models for $\eta(x)$ and $p(y \mid R = 1, x)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

2.3 Alternative factorization

In this paper, we develop novel estimators of μ_0 in the semiparametric model $\mathcal{M}$ defined by assumptions 1–3,

$$p(r, w; \varphi) = \pi(w; \eta, \gamma)^r \{1 - \pi(w; \eta, \gamma)\}^{1-r} \underbrace{p(y \mid u; \psi) p(z \mid u; \beta) p(u; \zeta)}_{p(w; \psi, \beta, \zeta)},$$

where γ is finite-dimensional and $(\eta, \psi, \beta, \zeta)$ is infinite-dimensional. The full data density is clearly separated from the nonresponse mechanism in the factorization above. This naturally encodes the exclusion restriction of assumption 2, which operates on the full data distribution, rather than on subpopulations defined by the nonresponse status. Furthermore, ψ and β are variationally independent, in the sense that any appropriate choice of ψ and β would result in a density in $\mathcal{M}$. Therefore, the models $p(y \mid u; \psi)$ and $p(z \mid u; \beta)$ can be specified separately without concerns about incompatibility. We show in the paper that estimation of μ_0 requires consistent estimation of at least a subset of the nuisance parameters (η, ψ, β). In modern studies, a broad collection of baseline covariates and operational features are usually recorded. When X is high-dimensional or contains several continuous components, nonparametric estimation is typically infeasible in the moderate sample sizes that are found in practice, as the data are too sparse due to the curse of dimensionality (Robins and Ritov, 1997). Thus, we are often forced to specify more stringent dimension-reducing models $\eta(x; \xi)$, $p(y \mid u; \psi)$ and $p(z \mid u; \beta)$. Although in principle these models could be made as flexible as allowed by the sample size, we focus on parametric specifications in this paper to ease exposition.

To mitigate the effects of model misspecifications, we develop novel semiparametric estimators of μ_0 which remain consistent and asymptotically normal in a union

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

model, where either one, but not necessarily both, of the following modeling assumptions hold:

(1) The models $\eta(x;\xi)$ and $p(z \mid u;\beta)$ are correctly specified such that $\eta(x) = \eta(x;\xi_0)$ and $p(z \mid u)$=$p(z \mid u;\beta_0)$ for some unknown finite-dimensional parameter vectors ξ_0 and β_0;

(2) The models $\eta(x;\xi)$ and $p(y \mid u;\psi)$ are correctly specified such that $\eta(x) = \eta(x;\xi_0)$ and $p(y \mid u) = p(y \mid u;\psi_0)$ for some unknown finite-dimensional parameter vectors ξ_0 and ψ_0;

Accordingly, we define the submodels $\mathcal{M}_1$, $\mathcal{M}_2$ of $\mathcal{M}$ which correspond to models (1), (2) respectively. Thus, the proposed estimators are doubly robust, as they deliver valid inferences in the union model $\cup_{j=1,2}\mathcal{M}_j$ where $\eta(x;\xi)$ (and hence the parametric extended propensity score model) is correctly specified, and either $p(z \mid u;\beta)$ or $p(y \mid u;\psi)$ is correctly specified. The proposed methodology requires correct specification of the model for $\eta(x)$, and therefore differs from the typical doubly robust inference when $\gamma = \gamma_0$ is known, where either the model for $\eta(x)$ or $p(y \mid R = 1, x)$, but not necessarily both, is correctly specified. It also differs from the doubly robust inference with instrumental variables proposed by Sun et al. (2018) when γ_0 is unknown, where the model for $p(z \mid u)$ is correctly specified, and either the model for $\eta(x)$ or $p(y \mid R = 1, x)$ is correctly specified. Specifying separate models for $\eta(x)$ and $p(y \mid R = 1, x)$ in a way that respects exclusion restriction is difficult, as they are inextricably entwined in the conditional density $p(y \mid x)$. On the other hand, because ψ and β are variationally independent under exclusion restriction, the proposed methodology provides the analyst with two genuine independent opportunities

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

to obtain valid inference.

2.4 Notation

Throughout, we use A^{T} to denote the transpose of a vector or matrix A, and $A_1 \otimes A_2$ to denote the Kronecker product of two vectors or matrices A_1 and A_2. Estimation and inference are based on an independent and identically distributed observed data sample $(O_1, ..., O_n)$, where $O = (R, RY, X)$. We denote the empirical measure as $\mathbb{P}_n$ so that empirical averages may be written as $n^{-1}\sum_{i=1}^{n}\{m(O_i)\} = \mathbb{P}_n\{m(O)\}$. For an arbitrary function $f(W)$ of the full data W, denote $f^{\dagger}(W) = \mathbb{E}\{f(W) \mid X\} + \mathbb{E}\{f(W) \mid Y, U\} - \mathbb{E}\{f(W) \mid U\}$.

2.5 An example with binary outcome and instrumental variable

We illustrate the key ideas with binary outcome and instrumental variable, $\mathcal{Y} = \mathcal{Z} = \{0, 1\}$. In this case, the exclusion restriction in assumption 2 is equivalent to $\mathrm{Cov}(Y, Z \mid U) = 0$, which can be captured by the moment restriction $\mathbb{E}[Y\{Z - p(Z = 1 \mid U)\} \mid U] = 0$. The main idea is that we can replace the moment function with its inverse propensity weighted form, to create the following observed data conditional moment restriction for the tilting parameter γ,

$$\mathbb{E}\left[\frac{R}{\pi(W;\eta,\gamma)}Y\{Z - p(Z = 1 \mid U)\}\middle| U\right] = 0, \tag{2.6}$$

which implies the following unconditional form,

$$\mathbb{E}\left[\frac{R}{\pi(W;\eta,\gamma)}d(U)Y\{Z - p(Z = 1 \mid U)\}\right] = 0, \tag{2.7}$$

where $d(u) \in \mathbb{R}^{p_\gamma}$ is a user-specified, vector-valued function with linearly independent elements. In principle, we can construct a semiparametric two-step estimator $\hat{\gamma}$ of γ_0

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

which solves the following empirical analogue of (2.7),

$$\mathbb{P}_n\left[\frac{R}{\pi(W;\hat{\eta},\gamma)}d(U)Y\{Z-\hat{p}(Z=1\mid U)\}\right]=0,$$

where $\hat{\eta}(x)$ is a nonparametric estimator of $\eta(x)$ based on (2.3) for each fixed value of γ, and $\hat{p}(Z=1\mid u)$ is a nonparametric estimator of $p(Z=1\mid u)$. We can then construct the following Hájek (1971) estimator of the outcome mean,

$$\hat{\mu}=\mathbb{P}_n\left\{\frac{RY}{\pi(W;\hat{\eta},\hat{\gamma})}\right\}\Big/\mathbb{P}_n\left\{\frac{R}{\pi(W;\hat{\eta},\hat{\gamma})}\right\}.$$

On the other hand, the zero conditional covariance restriction under assumption 2 can also be captured by the moment restriction $\mathbb{E}[\{Y-p(Y=1\mid U)\}Z\mid U]=0$. This yields the observed data conditional moment restriction

$$\mathbb{E}\left[\frac{R}{\pi(W;\eta,\gamma)}\{Y-p(Y=1\mid U)\}Z\middle|U\right]=0. \tag{2.8}$$

In contrast to the previous approach, $p(Y=1\mid u)$ cannot be directly estimated from observed data, but instead can be implicitly defined as the solution to the following inverse propensity weighted moment restriction,

$$\mathbb{E}\left[\frac{R}{\pi(W;\eta,\gamma)}\{Y-p(Y=1\mid U)\}\middle|U\right]=0. \tag{2.9}$$

A key observation is that (2.8) and (2.9) are functionally different. We can then construct a semiparametric two-step estimator of γ_0 which solves the following empirical moment condition for some vector-valued function $d(u)\in\mathbb{R}^{p_\gamma}$,

$$\mathbb{P}_n\left[\frac{R}{\pi(W;\hat{\eta},\gamma)}d(U)\{Y-\hat{p}(Y=1\mid U)\}Z\right]=0,$$

where $\hat{\eta}(x)$ is a nonparametric estimator of $\eta(x)$ based on (2.3), and $\hat{p}(Y=1\mid u)$ is a nonparametric estimator of $p(Y=1\mid u)$ based on (2.9), for each fixed value of γ. A Hájek (1971) estimator of the outcome mean can be constructed similarly.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

The preceding two approaches require estimation of $\eta(x)$, and either $p(z \mid u)$ or $p(y \mid u)$. In the presence of possibly high-dimensional X, we can perform inference under working parametric specifications $\eta(x;\xi)$, $p(z \mid u;\beta)$ and $p(y \mid u;\psi)$. We obtain $\hat{\theta} = (\hat{\xi}^{\mathrm{T}}, \hat{\beta}^{\mathrm{T}}, \hat{\psi}^{\mathrm{T}})^{\mathrm{T}}$, for each fixed value of γ, as the solution to the following unconditional empirical moments,

$$\begin{aligned} \mathbb{P}_n\left[\left\{\frac{R}{\pi(W;\xi,\gamma)} - 1\right\}\frac{\partial\eta(X;\xi)}{\partial\xi}\right] &= 0; \\ \mathbb{P}_n\left\{\frac{\partial \log p(Z \mid U;\beta)}{\partial\beta}\right\} &= 0; \\ \mathbb{P}_n\left[\left\{\frac{R}{\pi(W;\xi,\gamma)}\right\}\frac{\partial \log p(Y \mid U;\psi)}{\partial\psi}\right] &= 0. \end{aligned} \tag{2.10}$$

3. Doubly Robust Inference

3.1 Semiparametric theory

The construction of the Hájek (1971) estimators of μ_0 in the previous section involves preliminary estimation of the tilting parameter γ_0. As γ_0 encodes the degree of departure from ignorability, it is also sometimes of interest in its own right in missing data problems. To simplify the presentation, in this section we consider joint estimation of the unknown $(p_\gamma + 1)$-dimensional parameter $\phi_0 = (\mu_0, \gamma_0^{\mathrm{T}})^{\mathrm{T}}$. The construction of doubly robust estimators of ϕ_0 is often motivated by the form of the influence function in $\mathcal{M}$ (Robins and Rotnitzky, 2001; Chernozhukov et al., 2022). Specifically, any regular and asymptotically linear estimator $\hat{\phi}$ of ϕ_0 in $\mathcal{M}$ satisfies

$$n^{1/2}(\hat{\phi} - \phi_0) = n^{1/2}\mathbb{P}_n\{\Psi(O;\phi_0)\} + o_p(1),$$

where the i-th influence function $\Psi(O_i)$ represents the influence of the i-th observation on the estimator (Pfanzagl, 1982; Bickel et al., 1993; Newey, 1994; van der Laan and

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

Robins, 2003; Tsiatis, 2007). The estimation theory for ϕ_0 under general semiparametric models for the nonresponse mechanism has been previously developed (Robins et al., 2000; van der Laan and Robins, 2003). The result below, proved in Sun et al. (2018), provides an application of the general theory to model $\mathcal{M}$.

Result 1. The observed data influence function of any regular and asymptotically linear estimator of ϕ_0 in $\mathcal{M}$ is given by

$$\Psi(O;\phi_0,c,d) = -\left[\mathbb{E}\left\{\frac{\partial g(O;\phi,c,d)}{\partial\phi}\right\}\Big|_{\phi=\phi_0}\right]^{-1} g(O;\phi_0,c,d),$$

for some functions $c(W) \in \mathbb{R}$ and $d(W) \in \mathbb{R}^{p_\gamma}$ of the full data W, where

$$g(O;\phi,c,d) = \frac{Rq(W;\mu,c,d)}{\pi(W;\gamma)} + \left\{1 - \frac{R}{\pi(W;\gamma)}\right\}\mathbb{E}\left\{q(W;\mu,c,d)|R=0,X;\gamma\right\},$$

and

$$q(W;\mu,c,d) = \begin{Bmatrix} Y - \mu + c(W) - c^\dagger(W) \\ d(W) - d^\dagger(W) \end{Bmatrix} \in \mathbb{R}^{p_\gamma+1}.$$

The function $g(O;\phi,c,d)$ has the familiar inverse probability weighted form in missing data literature, augmented by an additional term involving the nonrespondents (Robins et al., 1994, 2000; van der Laan and Robins, 2003).

Example 2. In the special case with $\mathcal{Y} = \mathcal{Z} = \{0,1\}$, any function of the full data W can be expressed as $f(W) = f_0(U) + Yf_y(U) + Zf_z(U) + YZf_{yz}(U)$, where $\{f_0(U), f_y(U), f_z(U), f_{yz}(U)\}$ are functions of the same dimension as $f(W)$. It can be shown that

$$f(W) - f^\dagger(W) = f_{yz}(U)\{Y - p(Y=1 \mid U)\}\{Z - p(Z=1 \mid U)\}.$$

The example generalizes directly to discrete outcome and instrument taking values in $\mathcal{Y} = \{0,1,...,\ell_y\}$ and $\mathcal{Z} = \{0,1,...,\ell_z\}$, respectively. Let $v_1(Y) = \{I(Y =$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

$1), ..., I(Y = \ell_y)\}^{\mathrm{T}}$ and $v_2(Z) = \{I(Z = 1), ..., I(Z = \ell_z)\}^{\mathrm{T}}$, where $I(\cdot)$ is the indicator function. Then for any function of the full data W,

$$f(W) - f^{\dagger}(W) = f_{yz}(U)\{v_1(Y) - \mathbb{E}(v_1(Y) \mid U)\} \otimes \{v_2(Z) - \mathbb{E}(v_2(Z) \mid U)\},$$

for some conformable $f_{yz}(U)$.

Example 3. If at least one of Y and Z is continuous, inspired by the discrete case, we can simply discretize the continuous variables or use their moments. For example, if both Y and Z are continuous, we can consider the vectors $v_1(Y) = (Y, Y^2, ..., Y^{p_y})^{\mathrm{T}}$ and $v_2(Z) = (Z, Z^2, ..., Z^{p_z})^{\mathrm{T}}$, for some positive integers p_y and p_z.

3.2 Influence function-based doubly robust estimators

Following Robins and Rotnitzky (2001) and Chernozhukov et al. (2022), we can use $g(O; \phi, c, d)$ as a moment function to estimate ϕ_0. Evaluation of $g(O; \phi, c, d)$ relies on $\{\eta(x), p(z \mid u), p(y \mid u)\}$, which are directly targeted by the proposed approach. In particular, we note that the conditional outcome density among non-respondents can be expressed as

$$p(y \mid R = 0, x; \gamma) = \frac{\{1 - \pi(y, x; \gamma)\}p(y \mid u)}{1 - \int_{\mathcal{Y}} \pi(t, x; \gamma)p(t \mid u)d\nu(t)}, \tag{3.1}$$

which differs from (2.4) in terms of parameterization. When the baseline covariates include multiple continuous components, $\{\eta(x), p(z \mid u), p(y \mid u)\}$ can be estimated under user-specified, dimension-reducing parametric specifications $\{\eta(x; \xi), p(z \mid u; \beta), p(y \mid u; \psi)\}$, which allows for simpler conditions for asymptotic normality. Let $m(O; \gamma, \theta)$ denote the stacked moment functions in (2.10) for the parameter $\theta = (\xi^{\mathrm{T}}, \beta^{\mathrm{T}}, \psi^{\mathrm{T}})^{\mathrm{T}}$. The proposed influence function-based estimator $(\hat{\phi}^{\mathrm{T}}(c, d), \hat{\theta}^{\mathrm{T}})^{\mathrm{T}}$ then solves

$$\mathbb{P}_n\{g^{\mathrm{T}}(O; \phi, \theta, c, d), m(O; \gamma, \theta)\}^{\mathrm{T}} = 0.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

The asymptotic property of $\hat{\phi}(c,d)$ is given in the next proposition, where $\bar{\theta} = (\bar{\xi}^{\mathrm{T}}, \bar{\beta}^{\mathrm{T}}, \bar{\psi}^{\mathrm{T}})^{\mathrm{T}}$ denotes the probability limit of $\hat{\theta}$.

Proposition 1. *Under standard regularity conditions for moment estimation (Newey and McFadden, 1994), the estimator $\hat{\phi}(c,d)$ admits the following asymptotic expansion in the union model $\cup_{j=1,2}\mathcal{M}_j$,*

$$n^{1/2}(\hat{\phi}(c,d)-\phi_0) = -n^{1/2}\left[\mathbb{E}\left\{\frac{\partial}{\partial\phi}G(O;\phi,\bar{\theta},c,d)\right\}\Big|_{\phi=\phi_0}\right]^{-1}\mathbb{P}_n\{G(O;\phi_0,\bar{\theta},c,d)\}+o_p(1),$$

where

$$\begin{aligned}G(O;\phi,\bar{\theta},c,d) &= g(O;\phi,\bar{\theta},c,d)\\ &\quad-\mathbb{E}\left\{\frac{\partial}{\partial\theta}g(O;\phi_0,\theta,c,d)\right\}\Big|_{\theta=\bar{\theta}}\mathbb{E}\left\{\frac{\partial}{\partial\theta}m(O;\gamma_0,\theta)\right\}^{-1}\Big|_{\theta=\bar{\theta}}m(O;\gamma,\bar{\theta}).\end{aligned}$$

Furthermore, at the intersection submodel $\cap_{j=1,2}\mathcal{M}_j$ where all the working parametric models are correctly specified, $\hat{\phi}(c,d)$ admits the asymptotic expansion

$$n^{1/2}\{\hat{\phi}(c,d)-\phi_0\} = n^{1/2}\mathbb{P}_n\{\Psi(O;\phi_0,c,d)\}+o_p(1).$$

The first part of proposition 1 is due to the following double robustness property,

$$\mathbb{E}\{g(O;\phi_0,\bar{\xi},\bar{\beta},\bar{\psi},c,d)\}=0, \tag{3.2}$$

if either $\{\eta(x;\bar{\xi}),p(z\mid u;\bar{\beta})\}=\{\eta(x),p(z\mid u)\}$ or $\{\eta(x;\bar{\xi}),p(y\mid u;\bar{\psi})\}=\{\eta(x),p(y\mid u)\}$. Thus, deviations of $p(z\mid u;\bar{\beta})$ or $p(y\mid u;\bar{\psi})$ away from the truth has no global effect on the moment condition. The second part of proposition 1 states that estimation of the nuisance parameter θ has no first-order effect on the asymptotic expansion of $\hat{\phi}(c,d)$ at the intersection submodel $\cap_{j=1,2}\mathcal{M}_j$ where $\{\eta(x;\bar{\xi}),p(z\mid u;\bar{\beta}),p(y\mid u;\bar{\psi})\}=\{\eta(x),p(z\mid u),p(y\mid u)\}$. This is a general property of estimators that are constructed based on influence functions due to Neyman orthogonality (Chernozhukov et al., 2022),

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

which allows for simplification of the asymptotic variance formula. However, in general this simplification is lost as soon as one of the working models $p(z \mid u; \beta)$ or $p(y \mid u; \psi)$ is misspecified (Vermeulen and Vansteelandt, 2015).

3.3 Local efficiency

When both Y and Z are discrete, the efficient influence function in $\mathcal{M}$ is indexed by the optimal choice $(c, d) = (c^*, d^*)$, which is characterized in the supplementary material. At the intersection submodel $\cap_{j=1,2}\mathcal{M}_j$, the influence function of $\hat{\phi}(c, d)$ coincides with $\Psi(O; \phi_0, c, d)$, and hence the asymptotic variance of $\hat{\phi}(c^*, d^*)$ attains the semiparametric efficiency bound $\mathcal{V} = \mathbb{E}\{\Psi(O; \phi_0, c^*, d^*)\Psi(O; \phi_0, c^*, d^*)^{\mathrm{T}}\}$ (local efficiency). In practice, we can implement a doubly robust and locally efficient estimator of ϕ_0 based on a preliminary, consistent estimator of the optimal index functions, which we describe in the supplementary material. The results in Robins and Rotnitzky (2001) show that the efficiency bound in the union model $\cup_{j=1,2}\mathcal{M}_j$ coincides with the semiparametric efficiency bound $\mathcal{V}$ in $\mathcal{M}$. Therefore, $\hat{\phi}(c^*, d^*)$ also attains the efficiency bound in $\cup_{j=1,2}\mathcal{M}_j$ at the intersection submodel $\cap_{j=1,2}\mathcal{M}_j$. When at least one of Y and Z is continuous, the efficient influence function is typically not available in closed form. In this case, we can construct a doubly robust and approximately locally efficient estimator of ϕ_0, which is described in Sun et al. (2018).

3.4 A computationally simpler doubly robust estimator

The implementation of $\hat{\phi}(c, d)$ requires evaluation of the conditional outcome density among non-respondents given in (3.1). It is generally difficult to evaluate the integral in the denominator. For example, no closed form result for the logistic-normal inte-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

gral is known, although various approximations have been proposed in the literature (Crouch and Spiegelman, 1990). In this section, we propose a computationally simpler doubly robust estimator in the union model $\cup_{j=1,2}\mathcal{M}_j$ which avoids the numerical integration. The key observation is that double robustness (3.2) continues to hold for the inverse propensity weighted function $\tilde{g}(O;\phi,c,d) = Rq(W;\mu,c,d)/\pi(W;\gamma)$ which excludes the augmentation term involving non-respondents. Let $(\tilde{\phi}^{\mathrm{T}}(c,d), \hat{\theta}^{\mathrm{T}})^{\mathrm{T}}$ denote the joint solution to

$$\mathbb{P}_n\{\tilde{g}^{\mathrm{T}}(O;\phi,\theta,c,d), m(O;\gamma,\theta)\}^{\mathrm{T}} = 0.$$

Example 4. To illustrate the proposed computationally simpler doubly robust estimators, suppose $\mathcal{Y} = \mathcal{Z} = \{0,1\}$ and we set $c(W) = 0$. In this case, following the examples in section 2.5, estimation based on $\tilde{g}(O;\phi,c=0,d)$ can proceed in two stages. In the first stage, $(\tilde{\gamma}^{\mathrm{T}}, \tilde{\theta}^{\mathrm{T}})^{\mathrm{T}}$ solves (2.10) jointly with

$$\mathbb{P}_n\left[\frac{R}{\pi(W;\xi,\gamma)} d_{yz}(U)\{Y - p(Y=1 \mid U;\psi)\}\{Z - p(Z=1 \mid U;\beta)\}\right] = 0,$$

for some $d_{yz}(U) \in \mathbb{R}^{p_\gamma}$. Then the Hájek (1971) estimator $\tilde{\mu}$ of the outcome mean can be constructed with the estimated inverse propensity weights $\{R_i/\pi(W_i;\tilde{\xi},\tilde{\gamma}) : i = 1,...,n\}$.

The asymptotic property of $\tilde{\phi}(c,d)$ is summarized in the next proposition.

Proposition 2. *Under standard regularity conditions for moment estimation (Newey and McFadden, 1994), the estimator $\tilde{\phi}(c,d)$ admits the following asymptotic expansion in the union model $\cup_{j=1,2}\mathcal{M}_j$,*

$$n^{1/2}(\tilde{\phi}(c,d) - \phi_0) = -n^{1/2}\left[\mathbb{E}\left\{\frac{\partial}{\partial\phi}\tilde{G}(O;\phi,\bar{\theta},c,d)\right\}\Big|_{\phi=\phi_0}\right]^{-1} \mathbb{P}_n\{\tilde{G}(O;\phi_0,\bar{\theta},c,d)\} + o_p(1),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

where

$$\tilde{G}(O;\phi,\bar{\theta},c,d) = \tilde{g}(O;\phi,\bar{\theta},c,d) - \mathbb{E}\left\{\frac{\partial}{\partial\theta}\tilde{g}(O;\phi_0,\theta,c,d)\right\}\Big|_{\theta=\bar{\theta}} \mathbb{E}\left\{\frac{\partial}{\partial\theta}m(O;\gamma_0,\theta)\right\}^{-1}\Big|_{\theta=\bar{\theta}} m(O;\gamma,\bar{\theta}).$$

The asymptotic variance of $\tilde{\phi}(c,d)$ cannot attain the efficiency bound in $\mathcal{M}$ (and hence also in $\cup_{j=1,2}\mathcal{M}_j$) even at the intersection submodel $\cap_{j=1,2}\mathcal{M}_j$. Therefore, the estimator $\tilde{\phi}(c,d)$ is doubly robust but not locally efficient, and intuitively this is because it fails to incorporate the information from non-respondents when all the working models are correctly specified. Nonetheless, because we have already paid homage to the need for efficiency by using parametric models, the potential prize of attempting to attain local efficiency may not always be worth the chase in view of the additional computational demands. As the goal of this paper is to produce a statistically sound and practically useful method to tackle nonignorable missing data, we will focus on the estimator $\tilde{\phi}(c,d)$ in the simulation study and application.

4. Simulation Studies

In order to investigate the finite-sample properties of the doubly robust estimator proposed in Section 3, we perform Monte Carlo simulations involving identical and independently generated data $\{O_1,...,O_n\}$. The baseline covariate $U = (U_1,U_2)^{\mathrm{T}}$ is generated from a bivariate normal distribution $N(0,\Sigma)$, where the elements of Σ are $\sigma_1^2 = \sigma_2^2 = 1$ and $\sigma_{12} = 0.2$. Conditional on U, (R,Y,Z) is generated from the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

following generalized linear models consistent with assumptions 1–3,

$$\begin{aligned} Z \mid U &\sim \text{Bernoulli}\{p_1 = \text{expit}(1 + 2U_1 - U_2 - 0.8U_1U_2)\}, \\ Y \mid Z, U &\sim \text{Bernoulli}\{p_2 = \text{expit}(0.5 - 2U_1 + U_2)\}, \\ R \mid Y, X &\sim \text{Bernoulli}\{\pi = \text{expit}(2 - 3Z + 0.8U_1 + U_2 + \gamma Y)\}, \end{aligned}$$

where $\gamma = 2$. We implement the proposed doubly robust estimator $\tilde{\phi}(c, d) = (\tilde{\mu}, \tilde{\gamma})^{\mathrm{T}}$ with $c(w) = 0$, $d(w) = yz$ based on the models $\pi(w; \xi, \gamma) = \text{expit}\{(h_1^{\mathrm{T}}(u), z)^{\mathrm{T}}\xi + \gamma y\}$, $p(Z = 1 \mid u; \beta) = \text{expit}\{h_2(u)\beta\}$ and $p(Y = 1 \mid u; \psi) = \text{expit}\{h_3(u)\psi\}$ under the following five scenarios. The model for $\eta(x)$, $p(Z = 1 \mid u)$ or $p(Y = 1 \mid u)$ is misspecified if $h_j(u) = (1, u_1, u_1^2)$ for $j = 1$, 2 or 3, respectively.

(C1) Models for $\{\eta(x), p(Z = 1 \mid u), p(Y = 1 \mid u)\}$ are all correct.

(C2) Models for $\{\eta(x), p(Y = 1 \mid u)\}$ are correct, but misspecified for $p(Z = 1 \mid u)$.

(C3) Models for $\{\eta(x), p(Z = 1 \mid u)\}$ are correct, but misspecified for $p(Y = 1 \mid u)$.

(C4) Models for $\{p(Z = 1 \mid u), p(Y = 1 \mid u)\}$ are correct, but misspecified for $\eta(x)$.

(C5) Models for $\{\eta(x), p(Z = 1 \mid u), p(Y = 1 \mid u)\}$ are all misspecified.

To compare $\tilde{\mu}$ with the doubly robust estimator $\hat{\mu}_{\text{dr}}$ proposed by Sun et al. (2018), we specify the working parametric model $p(Y = 1 \mid R = 1, x; \psi_1, \lambda) = \text{expit}\{h_3(u)\psi_1 + \lambda z\}$ under (C1)–(C5) to evaluate its use in practice. In addition, we implement the complete-case estimator of the outcome mean $\hat{\mu}_{\text{cc}} = \mathbb{P}_n(RY)$, as well as the infeasible full-data estimator $\hat{\mu}_{\text{full}} = n^{-1}\sum_{i=1}^{n} Y_i$ as performance benchmark. For inference, we construct 95% Wald confidence intervals based on the sandwich estimator of asymptotic variance.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

The following remarks can be made based on the results of 1000 simulation replicates of sample size $n = 500$, 1000 or 5000 summarized in Table 1. The complete-case estimator $\hat{\mu}_{\text{cc}}$ exhibits severe bias and undercoverage. In agreement with theory, $\tilde{\mu}$ performs well in terms of bias and coverage cross scenarios (C1)–(C3), but is biased under (C4) and (C5) with misspecified model for $\eta(x)$. The estimator $\hat{\mu}_{\text{dr}}$ shows negligible bias and coverage proportion close to the nominal level under scenarios (C1) and (C3), but is biased under (C2) and (C5) with misspecified model for $p(Z = 1 \mid u)$. It also has small but noticeable bias under (C4). The relative efficiency of $\tilde{\mu}$ compared to the full-data estimator $\hat{\mu}_{\text{full}}$ is approximately 0.15 based on Monte Carlo variance at $n = 5000$. The supplementary material contains additional Monte Carlo simulation results under violations of the exclusion restriction in assumption 2, as well as with a continuous Y.

5. Illustration

To illustrate the proposed method, we analyze the household survey data on $n = 4997$ adults between the ages of 16 and 64 in Mochudi, Botswana, out of whom 4045 (81%) had complete information on HIV testing. The majority of those who did not have HIV test results refused to participate in the HIV testing component. The baseline covariates include participant gender (U_1) and age in years (U_2), while the candidate outcome instrument Z is a binary indicator of whether interviewer experience in years is in the top quartile. We implement the proposed estimator $\tilde{\phi}(c, d) = (\tilde{\mu}, \tilde{\gamma})^{\text{T}}$ with $c(w) = 0$, $d(w) = yz$ based on the following main effects generalized linear models with canonical links, $\pi(w; \xi, \gamma) = \text{expit}\{(1, z, u_1, u_2)\xi + \gamma y\}$, $p(Z = 1 \mid u; \beta) = \text{expit}\{(1, u_1, u_2)\beta\}$, and $p(Y = 1 \mid u; \psi) = \text{expit}\{(1, u_1, u_2)\psi\}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

Table 1: Summary of results for estimation of the outcome mean.

			(C1)		(C2)		(C3)		(C4)		(C5)	
			All correct		mis $p(z \mid u)$		mis $p(y \mid u)$		mis $\eta(x)$		All mis	
	$\hat{\mu}_{\text{cc}}$	$\hat{\mu}_{\text{full}}$	$\tilde{\mu}$	$\hat{\mu}_{\text{dr}}$	$\tilde{\mu}$	$\hat{\mu}_{\text{dr}}$	$\tilde{\mu}$	$\hat{\mu}_{\text{dr}}$	$\tilde{\mu}$	$\hat{\mu}_{\text{dr}}$	$\tilde{\mu}$	$\hat{\mu}_{\text{dr}}$
						$n = 500^{\dagger}$						
|Bias|	.136	.001	.004	.004	.002	.089	.006	.004	.042	.032	.093	.103
$\sqrt{\text{Var}}$	.024	.021	.055	.050	.055	.046	.057	.050	.049	.097	.047	.040
$\sqrt{\text{EVar}}$	.026	.022	.055	.080	.056	6.349	.064	.070	.059	.042	.048	.045
Cov95	.000	.957	.912	.943	.920	.948	.901	.943	.848	.847	.499	.386
						$n = 1000$						
|Bias|	.136	.000	.004	.001	.002	.088	.005	.001	.042	.017	.092	.101
$\sqrt{\text{Var}}$	.017	.016	.039	.036	.039	.032	.041	.036	.034	.034	.033	.028
$\sqrt{\text{EVar}}$	.018	.016	.037	.039	.038	1.813	.039	.039	.034	.029	.034	.032
Cov95	.000	.947	.916	.929	.931	.939	.920	.941	.750	.862	.219	.084
						$n = 5000$						
|Bias|	.136	.000	.001	.000	.001	.090	.002	.000	.044	.013	.093	.102
$\sqrt{\text{Var}}$	.008	.007	.017	.015	.017	.014	.018	.015	.015	.015	.015	.013
$\sqrt{\text{EVar}}$	.008	.007	.017	.016	.017	.066	.018	.016	.015	.013	.015	.014
Cov95	.000	.954	.939	.950	.943	.578	.933	.951	.169	.806	.000	.000

Note: †The results for $\hat{\mu}_{\text{dr}}$ excluded five simulation replicates due to convergence failure at $n = 500$. |Bias| and $\sqrt{\text{Var}}$ are the Monte Carlo absolute bias and standard deviation of the point estimates, $\sqrt{\text{EVar}}$ is the square root of the mean of the variance estimates and Cov95 is the coverage proportion of the 95% confidence intervals, based on 1000 repeated simulations. Zeros denote values smaller than .0005. The semiparametric efficiency bound for estimation of the outcome mean under the data generating mechanism of the simulation study is $\mathcal{V} \approx 1.2$ by Monte Carlo integration, with $\sqrt{\mathcal{V}/n} = .049, .035$ and .015 for $n = 500$, 1000 and 5000 respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

These parametric models are chosen due to their simplicity for illustration, although in principle they can be checked against the observed data using goodness-of-fit tests.

For comparison, we also implement the estimator $\hat{\mu}_{\mathrm{dr}}$ of Sun et al. (2018) based on the working model $p(Y = 1 \mid R = 1, x; \psi_1, \lambda) = \mathrm{expit}\{(1, u_1, u_2)\psi_1 + \lambda z\}$, as well as the standard complete-case estimator $\hat{\mu}_{\mathrm{cc}}$ and the inverse propensity weighted estimator $\hat{\mu}_{\mathrm{mar}}$ based on the missing at random propensity score model $\pi(w; \xi, \gamma = 0) = \mathrm{expit}\{(1, z, u_1, u_2)\xi\}$. The analysis results are summarized in Table 2. The point estimates of HIV seroprevalence for $\tilde{\mu}$ and $\hat{\mu}_{\mathrm{dr}}$ are similar and substantially higher than those for $\hat{\mu}_{\mathrm{cc}}$ and $\hat{\mu}_{\mathrm{mar}}$, although at the expense of higher variance. This difference in efficiency reflects genuine uncertainty about the underlying nonresponse mechanism, because $\hat{\mu}_{\mathrm{cc}}$ and $\hat{\mu}_{\mathrm{mar}}$ impose *a priori* restrictions on the parameter space of $(\xi^{\mathrm{T}}, \gamma)^{\mathrm{T}}$. The point estimate of $\tilde{\gamma}$ is -1.854 with 95% confidence interval $(-5.984, 2.277)$, which suggests that HIV-infected persons are less likely to participate in the HIV testing component of the survey, although this difference is not statistically significant at the 0.05 significance level.

6. Extension to longitudinal studies with repeated outcome measures

We follow the longitudinal study design in Vansteelandt et al. (2007), in which the full data at each study cycle $1 \leq t \leq T$ consists of $W_t = (Y_t, Z_t, U_t)$, where Y_t is an outcome, Z_t an instrumental variable such as interviewer experience, and U_t consists of other measured covariates. Suppose $X_t = (Z_t, U_t)$ is fully observed, but Y_t is subject to missingness due to some subjects missing some study cycles. Let R_t denote the occasion-specific missingness status. Thus, we only observe $O_t = (R_t, R_t Y_t, X_t)$ at study cycle t. Here the nonresponse patterns can be nonmonotone, as $R_t = 0$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

Table 2: Estimation of HIV prevalence based on a household survey data among adults in Mochudi, Botswana.

$\hat{\mu}_{\text{cc}}$	$\hat{\mu}_{\text{mar}}$	$\hat{\mu}_{\text{dr}}$	$\tilde{\mu}$
.214	.213	.284	.283
(.202, .227)	(.200, .225)	(.119, .449)	(.119, .447)

Note: Point estimates and 95% confidence intervals in parenthesis.

does not necessarily imply that $R_{t+1} = 0$. Let $\bar{O}_t = (O_1, ..., O_t)$ denote the observed history up to and including cycle t, with $\bar{O}_0 = \emptyset$. The nonresponse mechanism at each study cycle t is captured by the occasion-specific extended propensity score $\pi_t(w_t, \bar{o}_{t-1}) = p(R_t = 1 \mid W_t = w_t, \bar{O}_{t-1} = \bar{o}_{t-1})$.

Suppose our interest lie in estimating the mean of the outcome vector $Y = (Y_1, ..., Y_T)^{\text{T}}$, which we denote by $\mu_0 = (\mu_{1,0}, ..., \mu_{T,0})^{\text{T}}$. If the interviewers in the follow-up study are deployed randomly at the start of each cycle t, possibly within strata of all hitherto observed variables $(U_t, \bar{O}_{t-1})$, then it is plausible that the instrumental variable conditions $Z_t \not\perp\!\!\!\perp R_t \mid (U_t, \bar{O}_{t-1})$ and $Z_t \perp Y_t \mid (U_t, \bar{O}_{t-1})$ hold at each study cycle t. It follows that, analogous to the development in section 3, an occasion-specific doubly robust estimator of $\mu_{t,0}$ can be constructed for all $1 \leq t \leq T$. The resulting estimator of μ_0 is 2^T-multiply robust (Vansteelandt et al., 2007), as it remains consistent so long as one of two sets of parametric models is correctly specified at each study cycle t. This robustness against model misspecification is appealing

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

in view of the possibly high-dimensional data cumulatively observed over the study cycles.

7. Discussion

In closing, we acknowledge certain limitations of the proposed semiparametric method and outline avenues for future work. The proposed doubly robust estimators can be improved in terms of efficiency (Tan, 2006, 2010), bias (Vermeulen and Vansteelandt, 2015) and robustness (Han and Wang, 2013; Li et al., 2020). Our framework also opens the way to using various flexible nonparametric or machine learning methods to estimate the nuisance parameters (Chernozhukov et al., 2018, 2022). Although we have used the odds ratio, the proposed framework can in principle be extended to other measures of the residual association between the outcome of interest and nonresponse mechanism, for example on the multiplicative or additive scales. Lastly, nonignorable missing covariate data is also a long-standing problem in applied research, and methods for identification and inference abound. For example, Miao and Tchetgen Tchetgen (2018) used shadow variables, while Bartlett et al. (2014) and Yang et al. (2019) assume that the covariate missingness mechanism is independent of the outcome. It will be of interest to explore the use of instrumental variables in similar settings with missing covariate data.

Supplementary Material

The online Supplementary Material contains proofs of propositions 1 and 2, a further discussion on local efficiency and additional simulation results.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

Acknowledgments

Baoluo Sun's work is supported by the Ministry of Education, Singapore, under its Academic Research Fund Tier 1 (A-8000452-00-00). Wang Miao's research is supported by National Key R&D Program of China (2022YFA1008100). The authors would like to thank the anonymous referees, an Associate Editor and the Editor for their constructive comments that led to a much improved paper.

References

Ahn, H. and J. L. Powell (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *J. Economet. 58*(1-2), 3–29.

Bang, H. and J. M. Robins (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics 61*(4), 962–973.

Bärnighausen, T., J. Bor, S. Wandira-Kazibwe, and D. Canning (2011). Correcting hiv prevalence estimates for survey nonparticipation using heckman-type selection models. *Epidemiology 22*(1), 27–35.

Bartlett, J. W., J. R. Carpenter, K. Tilling, and S. Vansteelandt (2014). Improving upon the efficiency of complete case analysis when covariates are MNAR. *Biostatistics 15*(4), 719–730.

Bickel, P. J., C. A. Klaassen, P. J. Bickel, Y. Ritov, J. Klaassen, J. A. Wellner, and Y. Ritov (1993). *Efficient and Adaptive Estimation for Semiparametric Models*, Volume 4. Johns Hopkins University Press Baltimore.

Chen, H., Z. Geng, and X.-H. Zhou (2009). Identifiability and estimation of causal effects in randomized trials with noncompliance and completely nonignorable missing data. *Biometrics 65*(3), 675–682.

Chen, H. Y. (2007). A semiparametric odds ratio model for measuring association. *Biometrics 63*(2), 413–421.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018, 01). Double/debiased machine learning for treatment and structural parameters. *Economet. J. 21*(1), C1–C68.

Chernozhukov, V., J. C. Escanciano, H. Ichimura, W. K. Newey, and J. M. Robins (2022). Locally robust semiparametric estimation. *Econometrica 90*(4), 1501–1535.

Crouch, E. A. and D. Spiegelman (1990). The evaluation of integrals of the form $\int_{-\infty}^{+\infty} f(t)\exp(-t^2)dt$: Application to logistic-normal models. *J. Am. Statist. Assoc. 85*(410), 464–469.

Das, M., W. K. Newey, and F. Vella (2003). Nonparametric estimation of sample selection models. *The Review of Economic Studies 70*(1), 33–58.

d'Haultfoeuille, X. (2010). A new instrumental method for dealing with endogenous selection. *J. Economet. 154*(1), 1–15.

Hájek, J. (1971). Comment on a paper by d. basu. In V. P. Godambe and D. A. Sprott (Eds.), *Foundations of statistical inference*, pp. 236. Toronto: Holt, Rinehart and Winston.

Han, P. and L. Wang (2013). Estimation with missing data: beyond double robustness. *Biometrika 100*(2), 417–430.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica 42*(4), 679–694.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Kim, J. K. and J. Shao (2021). *Statistical Methods for Handling Incomplete Data*. Chapman and Hall/CRC.

Kim, J. K. and C. L. Yu (2011). A semiparametric estimation of mean functionals with nonignorable missing data. *J. Am. Statist. Assoc. 106*(493), 157–165.

Lepkowski, J. M., M. P. Couper, and R. M. Groves (2002). Nonresponse in the second wave of longitudinal household surveys, international conference in survey nonresponse. In *International conference in survey nonresponse*, pp. 259–274. New York: Wiley;.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

Li, W., Y. Gu, and L. Liu (2020). Demystifying a class of multiply robust estimators. *Biometrika 107*(4), 919–933.

Li, W., W. Miao, and E. J. Tchetgen Tchetgen (2023). Non-parametric inference about mean functionals of non-ignorable non-response data without identifying the joint distribution. *J. R. Stat. Soc.* B *85*(3), 913–935.

Liang, K.-Y. and J. Qin (2000). Regression analysis under non-standard situations: a pairwise pseudolikelihood approach. *J. R. Stat. Soc.* B *62*(4), 773–786.

Malinsky, D., I. Shpitser, and E. J. Tchetgen Tchetgen (2022). Semiparametric inference for nonmonotone missing-not-at-random data: the no self-censoring model. *J. Am. Statist. Assoc. 117*(539), 1415–1423.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review 80*(2), 319–323.

Marden, J. R., L. Wang, E. J. Tchetgen Tchetgen, S. Walter, M. M. Glymour, and K. E. Wirth (2018). Implementation of instrumental variable bounds for data missing not at random. *Epidemiology 29*(3), 364–368.

Miao, W., L. Liu, Y. Li, E. J. Tchetgen Tchetgen, and Z. Geng (2024). Identification and semiparametric efficiency theory of nonignorable missing data with a shadow variable. *ACM / IMS J. Data Sci. 1*(2), 1–23.

Miao, W. and E. J. Tchetgen Tchetgen (2016). On varieties of doubly robust estimators under missingness not at random with a shadow variable. *Biometrika 103*(2), 475–482.

Miao, W. and E. J. Tchetgen Tchetgen (2018). Identification and inference with nonignorable missing covariate data. *Statist. Sinica 28*(4), 2049–2067.

Molenberghs, G., G. Fitzmaurice, M. G. Kenward, A. Tsiatis, and G. Verbeke (2014). *Handbook of Missing Data Methodology*. CRC Press.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica: Journal of the Econometric Society 62*(6), 1349–1382.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, Volume 4, pp. 2111–2245. Elsevier.

Nicoletti, C. and F. Peracchi (2005). Survey response and survey characteristics: microlevel evidence from the European Community Household Panel. *J. R. Stat. Soc.* A *168*(4), 763–781.

Pfanzagl, J. (1982). *Contributions to a General Asymptotic Statistical Theory*. Springer.

Powell, J. L. (1994). Estimation of semiparametric models. In *Handbook of Econometrics*, Volume 4, pp. 2443–2521. Elsevier.

Riddles, M. K., J. K. Kim, and J. Im (2016). A propensity-score-adjustment method for nonignorable nonresponse. *Journal of Survey Statistics and Methodology 4*(2), 215–245.

Robins, J. M. and Y. Ritov (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statist. Med. 16*(3), 285–319.

Robins, J. M. and A. Rotnitzky (2001). Comment on the bickel and kwon article,"inference for semiparametric models: Some questions and an answer". *Statist. Sinica 11*(4), 920–936.

Robins, J. M., A. Rotnitzky, and D. O. Scharfstein (2000). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In M. E. Halloran and D. Berry (Eds.), *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, New York, NY, pp. 1–94. Springer New York.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Am. Statist. Assoc. 89*(427), 846–866.

Rotnitzky, A., J. M. Robins, and D. O. Scharfstein (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *J. Am. Statist. Assoc. 93*(444), 1321–1339.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

Rubin, D. B. (1976). Inference and missing data. *Biometrika 63*(3), 581–592.

Rubin, D. B. and R. J. Little (2019). *Statistical Analysis with Missing Data*. John Wiley & Sons.

Scharfstein, D. O., A. Rotnitzky, and J. M. Robins (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Am. Statist. Assoc. 94*(448), 1096–1120.

Schräpler, J.-P. (2004). Respondent behavior in panel studies: A case study for income nonresponse by means of the german socio-economic panel (SOEP). *Sociological Methods & Research 33*(1), 118–156.

Seaman, S. R. and S. Vansteelandt (2018). Introduction to double robust methods for incomplete data. *Statistical science 33*(2), 184–197.

Shao, J. and L. Wang (2016). Semiparametric inverse propensity weighting for nonignorable missing data. *Biometrika 103*(1), 175–187.

Sun, B., L. Liu, W. Miao, K. Wirth, J. Robins, and E. J. Tchetgen Tchetgen (2018). Semiparametric estimation with data missing not at random using an instrumental variable. *Statist. Sinica 28*(4), 1965–1983.

Tan, Z. (2006). A distributional approach for causal inference using propensity scores. *J. Am. Statist. Assoc. 101*(476), 1619–1637.

Tan, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika 97*(3), 661–682.

Tang, G., R. J. Little, and T. E. Raghunathan (2003). Analysis of multivariate missing data with nonignorable nonresponse. *Biometrika 90*(4), 747–764.

Tchetgen Tchetgen, E. J., J. M. Robins, and A. Rotnitzky (2010). On doubly robust estimation in a semiparametric odds ratio model. *Biometrika 97*(1), 171–180.

Tchetgen Tchetgen, E. J. and K. E. Wirth (2017). A general instrumental variable framework for regression analysis with outcome missing not at random. *Biometrics 73*(4), 1123–1131.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0383

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

van der Laan, M. J. and J. M. Robins (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer.

Vansteelandt, S., A. Rotnitzky, and J. Robins (2007). Estimation of regression models for the mean of repeated outcomes under nonignorable nonmonotone nonresponse. *Biometrika 94*(4), 841–860.

Vermeulen, K. and S. Vansteelandt (2015). Bias-reduced doubly robust estimation. *J. Am. Statist. Assoc. 110*(511), 1024–1036.

Wang, S., J. Shao, and J. K. Kim (2014). An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statist. Sinica 24*(3), 1097–1116.

Yang, S., L. Wang, and P. Ding (2019). Causal inference with confounders missing not at random. *Biometrika 106*(4), 875–888.

Zhao, J. and Y. Ma (2022). A versatile estimation procedure without estimating the nonignorable missingness mechanism. *J. Am. Statist. Assoc. 117*(540), 1916–1930.

Zhao, J. and J. Shao (2015). Semiparametric pseudo-likelihoods in generalized linear models with nonignorable missing data. *J. Am. Statist. Assoc. 110*(512), 1577–1590.

Department of Statistics and Data Science, National University of Singapore, Singapore.

E-mail: stasb@nus.edu.sg

Department of Probability and Statistics, Peking University, China.

E-mail: mwfy@pku.edu.cn

Department of Statistics, University of Colombo, Sri Lanka.

E-mail: deshanee@stat.cmb.ac.lk