Biometrika (2016), *pp.* 1–13
Printed in Great Britain
doi: 10.1093/biomet/asv071

Semiparametric inverse propensity weighting for nonignorable missing data

JUN SHAO

School of Statistics, East China Normal University, Shanghai 200241, China
shao@stat.wisc.edu

AND LEI WANG

Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin 53706, U.S.A.
leiwang.stat@gmail.com

SUMMARY

To estimate unknown population parameters based on data having nonignorable missing values with a semiparametric exponential tilting propensity, Kim & Yu (2011) assumed that the tilting parameter is known or can be estimated from external data, in order to avoid the identifiability issue. To remove this serious limitation on the methodology, we use an instrument, i.e., a covariate related to the study variable but unrelated to the missing data propensity, to construct some estimating equations. Because these estimating equations are semiparametric, we profile the nonparametric component using a kernel-type estimator and then estimate the tilting parameter based on the profiled estimating equations and the generalized method of moments. Once the tilting parameter is estimated, so is the propensity, and then other population parameters can be estimated using the inverse propensity weighting approach. Consistency and asymptotic normality of the proposed estimators are established. The finite-sample performance of the estimators is studied through simulation, and a real-data example is also presented.

Some key words: Exponential tilting; Generalized method of moments; Identifiability; Instrumental variable; Kernel regression; Nonignorable nonresponse.

1. INTRODUCTION

Many well-established methods are available to handle missing data that are ignorable or missing at random in the sense that the propensity of missing data depends only on the observed values. In many applications, however, the missing data are nonignorable, i.e., the propensity depends not only on observed data but also on unobserved data. Nonignorable missingness presents a challenge in applications, as the treatment of incomplete data requires assumptions that are hard to verify due to nonignorable missingness, and estimation may be seriously biased without proper treatment.

We assume that y, a univariate outcome or response of interest, is subject to nonresponse and that x, a vector of auxiliary variables, is observed for the entire sample. Let δ be the response status indicator for y. The conditional probability $\pi(y, x) = \mathrm{pr}(\delta = 1 \mid y, x)$ is called the propensity of missing data. In the nonignorable case, the propensity depends on y regardless of whether y is observed or missing. The strongest assumption is to impose parametric models on both $\pi(y, x)$ and the probability density of y given x, $f(y \mid x)$; see Molenberghs & Kenward (2007).

The fully parametric approach is sensitive to failure of the assumed models (Little, 1985). On the other hand, when both $\pi(y, x)$ and $f(y \mid x)$ are fully unspecified or nonparametric, the distribution of (y, δ) given x, $\pi(y, x) f(y \mid x)$, is nonidentifiable (Robins & Ritov, 1997), so that valid statistical analysis may not be possible. More reasonable approaches are semiparametric. For example, Tang et al. (2003) proposed a pseudolikelihood method, assuming a parametric model for the density $f(y \mid x)$ but allowing the propensity of missing data to be unspecified. In this paper we focus on the approach proposed by Qin et al. (2002) and followed up by Wang et al. (2014), which imposes a parametric model on $\pi(y, x)$ but allows $f(y \mid x)$ to be unspecified.

Qin et al. (2002) and Wang et al. (2014) assumed a purely parametric model on the propensity $\pi(y, x)$. Because any model on $\pi(y, x)$ is difficult to verify under nonignorable missingness, it is desirable to have a model assumption that is as weak as possible. To this end, Kim & Yu (2011) considered the following exponential tilting model for the propensity:

$$\pi(y, x) = [1 + \exp\{g(x) + \gamma y\}]^{-1}, \tag{1}$$

where γ is an unknown tilting parameter and $g(x)$ is a completely unspecified function of x. Note that (1) is a semiparametric model weaker than the parametric model assumed by Qin et al. (2002). However, to estimate unknown characteristics of $f(y \mid x)$ and the propensity $\pi(y, x)$, Kim & Yu (2011) assumed that the tilting parameter γ in (1) is known or can be estimated using external data. Zhao et al. (2013) and Tang et al. (2014) refined the approach of Kim & Yu (2011), but they still require a known γ or a γ estimated from external data. While model (1) is a great improvement over a parametric model for the propensity, the requirement of a γ that is known or estimable from external data seriously limits its application.

Under model (1), without any further assumptions, g and γ are not identifiable when they are both unknown; see Example 1. In this paper, we show that all unknown parameters can be identified and consistently estimated if we can find some part of x that is not involved in (1), i.e., if $x = (u, z)$ and

$$\pi(y, x) = [1 + \exp\{g(u) + \gamma y\}]^{-1}, \tag{2}$$

where g is still an unknown and unspecified function of u. Note that z does not appear in (2) but has to be a useful covariate for y. Such a covariate, termed a nonresponse instrument by Wang et al. (2014), allows us to identify and estimate all unknown parameters without requiring external data. In fact, our approach includes the use of external data as a special case, since we can define an instrument z having two categories, the original sample and the external dataset. Furthermore, our approach can be extended to the case where γy in (2) is replaced by $h_\gamma(y)$, a parametric function of y with an unknown parameter vector γ.

Different approaches based on instrumental variables exist. For example, one could utilize an instrument that is related to nonresponse propensity but not related to y given other covariates (Yang et al., 2014).

The key ingredients of our approach are the construction of some estimating equations and the profiling of the nonparametric component g. Once g is profiled and estimated by a kernel-type estimator, we can estimate γ using the profiled estimating equations and the generalized method of moments. We establish the consistency and asymptotic normality of the proposed estimators, and illustrate their performance using simulations and a real-data example. The technical details are given in the Supplementary Material.

2. Methodology

2·1. *Identifiability*

Let $\{(x_i, y_i, \delta_i) : i = 1, \ldots, n\}$ be an independent and identically distributed sample from (x, y, δ), where the covariate vector x_i is fully observed and the response y_i is observed if and only if $\delta_i = 1$.

Identifiability is challenging when there are nonignorable missing data (Fitzmaurice et al., 1995). We say that a population is identifiable if two different populations, $\mathcal{P}$ and $\mathcal{P}'$, do not give the same $\pi(y, x) f(y \mid x)$, where $f(y \mid x)$ is the density of y conditioned on x and $\pi(y, x)$ is the propensity of missing data. If a population is not identifiable, then some characteristics of the population cannot be adequately estimated.

Under model (1), with unknown g and γ and no further conditions on the propensity, the following example shows that the population is not identifiable.

Example 1. Suppose that (1) holds, with $\gamma \neq 0$ and $g(x)$ unspecified, and let $f(y \mid x)$ be normal with mean $\mu(x)$ and standard deviation $\sigma(x)$, both of which are unspecified functions of x so that $f(y \mid x)$ is semiparametric. If there exist two different sets $\{g(x), \mu(x), \sigma(x), \gamma\}$ and $\{g'(x), \mu'(x), \sigma'(x), \gamma'\}$ such that for all x and y,

$$\frac{\exp[-\{y - \mu(x)\}^2/\{2\sigma^2(x)\}]}{\sigma(x)[1 + \exp\{g(x) + \gamma y\}]} = \frac{\exp[-\{y - \mu'(x)\}^2/\{2\sigma'^2(x)\}]}{\sigma'(x)[1 + \exp\{g'(x) + \gamma' y\}]},$$

then the population is not identifiable. It can be verified that the above identity holds when $\gamma' = -\gamma$, $g'(x) = -g(x)$, $\sigma'(x) = \sigma(x)$, $\mu'(x) = \mu(x) - \sigma^2(x)\gamma$ and $g(x) = \sigma^2(x)\gamma^2/2 - \mu(x)\gamma$ for completely unspecified $\mu(x)$ and $\sigma(x)$.

The situation is different if γ is known. Let $\pi(y, x, g, \gamma) = 1/[1 + \exp\{g(x) + \gamma y\}]$. Under model (1),

$$E\left\{\frac{\delta}{\pi(y, x, g, \gamma)} - 1 \,\middle|\, x\right\} = 0, \tag{3}$$

which is equivalent to

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

With a known $\gamma = \gamma_0$, Kim & Yu (2011) proposed the following kernel regression estimator for $\exp\{g(x)\}$:

$$\exp\{\hat{g}(x)\} = \frac{\sum_{i=1}^n (1 - \delta_i) K_h(x - x_i)}{\sum_{i=1}^n \delta_i \exp(\gamma_0 y_i) K_h(x - x_i)}, \tag{4}$$

where $K_h(\cdot) = h^{-1} K(\cdot/h)$, with $K(\cdot)$ being a symmetric kernel function and h a bandwidth. Based on (4), one can apply inverse propensity weighting to estimate $E\{\psi(y, x)\}$ with any known function ψ by

$$\frac{1}{n} \sum_{i=1}^n \frac{\delta_i \psi(y_i, x_i)}{\hat{\pi}(y_i, x_i)}, \quad \hat{\pi}(y, x) = \frac{1}{1 + \exp\{\hat{g}(x) + \gamma_0 y\}}.$$

This method fails if γ is unknown. When both g and γ are unknown, (3) defines one estimating equation and cannot be used to estimate two unknown quantities.

2·2. *Instrumental estimating equation and profiling*

To solve the identifiability problem without requiring a known γ or external data to estimate γ, some extra assumptions are needed to create more estimating equations in addition to (3). As discussed in § 1, we assume that there is an instrument z satisfying (2).

The existence of this instrument can be used to create more estimating equations. To illustrate the idea, consider the special case where $g(u)$ in (2) is just an unknown constant λ; that is,

$$\pi(y, x) = \pi(y, \lambda, \gamma) = \frac{1}{1 + \exp(\lambda + \gamma y)}$$

is a function of y only, but the parameters λ and γ are unknown. In this case, (3) reduces to

$$E\left\{\frac{\delta}{\pi(y, \lambda, \gamma)} - 1\right\} = 0. \tag{5}$$

There are two unknown parameters, λ and γ, in the single equation (5), so we cannot estimate them without help. Notice that $\pi(y, \lambda, \gamma)$ does not depend on the covariate x. Hence $z = x$ satisfies one of the two basic requirements for being an instrument; the other requirement is that x must be related to y. Suppose that x is a binary covariate. It can be shown that

$$E\left[x\left\{\frac{\delta}{\pi(y, \lambda, \gamma)} - 1\right\}\right] = 0. \tag{6}$$

The two equations (5) and (6) can be used to estimate the two parameters λ and γ if the equations are not linearly related. The left-hand side of (6) equals

$$E\left(E\left[x\left\{\frac{\delta}{\pi(y, \lambda, \gamma)} - 1\right\}\,\middle|\, y\right]\right) = E\left[\mathrm{pr}(x = 1 \mid y) E\left\{\frac{\delta}{\pi(y, \lambda, \gamma)} - 1\,\middle|\, y\right\}\right],$$

since x and δ are conditionally independent given y. If x is independent of y, then $\mathrm{pr}(x = 1 \mid y)$ is constant, so (6) is the same as (5). If x is an instrument, then $\mathrm{pr}(x = 1 \mid y)$ depends on y. Hence (5) and (6) are two different equations and can be used to estimate λ and γ.

We now consider the general situation. Under model (2), equation (3) holds with x replaced by u. Let z be a discrete instrument taking values $l = 1, \ldots, L$. For each l, let

$$m_l(y, u, \delta, g, \gamma) = I(z = l)\left\{\frac{\delta}{\pi(y, u, g, \gamma)} - 1\right\}, \tag{7}$$

where $I(\cdot)$ is the indicator function, and let $m(y, u, \delta, g, \gamma)$ be the L-dimensional vector whose lth component is m_l. Then

$$E\{m(y, u, \delta, g, \gamma) \mid u\} = 0. \tag{8}$$

This defines L estimating equations; since they are constructed using the instrumental variable z, we call them instrumental estimating equations. When $L = 1$, i.e., if z is a constant or there is no instrumental variable, (8) reduces to (3) and is insufficient for estimating unknown g and γ.

To construct a sample version of the estimating equation (8), we need to overcome the difficulty arising from g in (8) being nonparametric, i.e., that the instrumental estimating equations are semiparametric equations. We use the idea of profiling. For every fixed γ in the parameter

space, we first use the sum of L equations in (8) to obtain the following kernel estimate similar to (4):

$$\exp\{\hat{g}_\gamma(u)\} = \frac{\sum_{i=1}^n (1-\delta_i) K_h(u-u_i)}{\sum_{i=1}^n \delta_i \exp(\gamma y_i) K_h(u-u_i)}, \tag{9}$$

where the bandwidth h may depend on l. The difference between (4) and (9) is that γ_0 in (4) is known whereas γ in (9) is an unknown parameter value, so that $\hat{g}_\gamma$ depends on γ; $\hat{g}_\gamma$ is not an estimator of g unless γ is the true parameter value, but $\hat{g}_\gamma$ is useful for profiling. Since profiling uses the sum of L equations in (8), there are $L-1$ equations left for estimating the unknown γ; hence we need $L \geqslant 2$. Once we have $\hat{g}_\gamma(u)$, we can obtain $L-1$ profile equations using (8) with g replaced by $\hat{g}_\gamma(u)$.

If γy in (2) is extended to a parametric function $h_\gamma(y)$ with a q-dimensional unknown vector γ, then the previous idea still works as long as $L \geqslant q+1$, i.e., provided the instrument has at least $q+1$ categories. If a continuous covariate z is suitable as an instrument, then we can apply the proposed approach by discretizing z or replacing equation (7) with moment-based estimating equations, for example replacing the indicator $I(z=l)$ in (7) with the lth moment of z $(l=1,\ldots,L)$.

2·3. *Generalized method of moments and inverse propensity weighting*

A sample version of the instrumental estimating equation (8) is

$$\frac{1}{n}\sum_{i=1}^n m_l(y_i, u_i, \delta_i, \hat{g}_\gamma, \gamma) = 0 \quad (l=1,\ldots,L), \tag{10}$$

where the function m_l is defined by (8) and $\hat{g}_\gamma(u)$ is given in (9). When L is greater than 1 plus the dimension of γ, (10) cannot be solved because it over-identifies; in that case we employ the two-step generalized method of moments (Hansen, 1982). Let $\bar{M}_n(\hat{g}_\gamma, \gamma)$ be the L-dimensional vector whose lth component is the left-hand side of (10). The first-step generalized moment estimator of γ is

$$\hat{\gamma}^{(1)} = \arg\min_{\gamma\in\Upsilon} \bar{M}_n(\hat{g}_\gamma, \gamma)^{\mathrm{T}} \bar{M}_n(\hat{g}_\gamma, \gamma),$$

where Υ is the parameter space for γ and a superscript T denotes the transpose of a vector. Let $\hat{W}_n$ be the inverse matrix of the $L \times L$ matrix with (l, l')th element

$$\frac{1}{n}\sum_{i=1}^n m_l\big(y_i, u_i, \delta_i, \hat{g}_{\hat{\gamma}^{(1)}}, \hat{\gamma}^{(1)}\big) m_{l'}\big(y_i, u_i, \delta_i, \hat{g}_{\hat{\gamma}^{(1)}}, \hat{\gamma}^{(1)}\big),$$

which is assumed to converge in probability to a positive-definite matrix W^{-1} for sufficiently large n. The efficient second-step generalized moment estimator of γ is

$$\hat{\gamma} = \arg\min_{\gamma\in\Upsilon} \bar{Q}_n(\hat{g}_\gamma, \gamma), \tag{11}$$

with $\bar{Q}_n(\hat{g}_\gamma, \gamma) = \bar{M}_n(\hat{g}_\gamma, \gamma)^{\mathrm{T}} \hat{W}_n \bar{M}_n(\hat{g}_\gamma, \gamma)$.

Once $\hat{\gamma}$ in (11) is obtained, we estimate $g(u)$ by $\hat{g}_{\hat{\gamma}}(u)$. Estimators of unknown quantities in $f(y \mid x)$ or the marginal of y can be obtained using inverse propensity weighting with the

estimated propensity as the weight function. To estimate $\mu = E(y)$, the mean of y, we use

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i y_i}{\pi(y_i, u_i, \hat{g}_{\hat{\gamma}}, \hat{\gamma})}. \tag{12}$$

If we are interested in a parameter vector θ which is a unique solution to $E\{\eta(y, x, \theta)\} = 0$, then we estimate θ by a solution to

$$\sum_{i=1}^{n} \frac{\delta_i \eta(y_i, x_i, \theta)}{\pi(y_i, u_i, \hat{g}_{\hat{\gamma}}, \hat{\gamma})} = 0.$$

For example, to estimate the distribution function of y at a fixed value t, we use $\eta(y, x, \theta) = I(y \leqslant t) - \theta$; to estimate θ under a linear model $E(y \mid x) = x^{\mathrm{T}}\theta$, we use $\eta(y, x, \theta) = x(y - x^{\mathrm{T}}\theta)$. A nonparametric estimator of $E(y \mid x)$ can be similarly obtained.

3. THEORY

In this section, we study the consistency and asymptotic normality of the semiparametric two-step generalized moment estimator $\hat{\gamma}$ in (11) and the estimator $\hat{\mu}$ given by (12). Although the kernel estimator $\hat{g}_\gamma$ has a slower than $n^{1/2}$ convergence rate, the semiparametric two-step generalized moment estimator can be shown to be $n^{1/2}$-consistent and asymptotically normal under the following conditions, as in Greblicki et al. (1984), Newey (1994), Newey & McFadden (1994) and Khan & Powell (2001).

Condition 1. The kernel $K(u)$ has bounded derivatives of order d, satisfies $\int K(u)\,\mathrm{d}u = 1$, and has zero moments of order up to $m - 1$ and nonzero mth-order moment.

Condition 2. The true function of $g(u)$ is continuously differentiable and bounded on an open set containing the support of u.

Condition 3. The moment $E\{\exp(4\gamma y)\}$ is finite and the function $E\{\exp(4\gamma y) \mid u\} f(u)$ is bounded, where $f(u)$ is the marginal density of u.

Condition 4. The bandwidth $h = h_n$ is such that $h_n \to 0$, $nh_n^p \to \infty$, $n^{1/2}h_n^{p+2d}/\log n \to \infty$, and $nh_n^{2m} \to 0$ as $n \to \infty$, where p is the dimension of u.

For any matrix A, define $\|A\|^2 = \mathrm{tr}(A^{\mathrm{T}}A)$. Throughout, γ_0 denotes the true unknown value of the tilting parameter γ and

$$\exp\{g_\gamma(u)\} = \frac{E\{(1-\delta) \mid u\}}{E\{\delta \exp(\gamma y) \mid u\}} \tag{13}$$

so that $g(u) = g_{\gamma_0}(u)$.

THEOREM 1. *Assume that Conditions 1–4 hold, that γ_0 is the unique solution to $m_0(g_\gamma, \gamma) = E\{m(y, u, \delta, g_\gamma, \gamma)\}$ and that $\sup_{\gamma\in\Upsilon} \|m_0(g_\gamma, \gamma)\| < \infty$. Then, as $n \to \infty$, $\hat{\gamma} \to \gamma_0$ in probability.*

Let $\nabla_\gamma(\cdot)$ and $\nabla_{\gamma\gamma}(\cdot)$ denote the first- and second-order derivatives with respect to γ. Under the assumptions in Theorem 1, $\nabla_\gamma \bar{Q}_n(\hat{g}_{\hat{\gamma}}, \hat{\gamma}) = 0$ with probability approaching 1. Expanding

$\nabla_\gamma \bar{Q}_n(\hat{g}_{\hat{\gamma}}, \hat{\gamma})$ about γ_0 gives

$$\nabla_\gamma \bar{Q}_n(\hat{g}_{\hat{\gamma}}, \hat{\gamma}) = \nabla_\gamma \bar{Q}_n(\hat{g}_{\gamma_0}, \gamma_0) + \nabla_{\gamma\gamma} \bar{Q}_n(\hat{g}_{\tilde{\gamma}}, \tilde{\gamma})(\hat{\gamma} - \gamma_0),$$

where $\tilde{\gamma}$ is between $\hat{\gamma}$ and γ_0. Therefore

$$\begin{aligned} n^{1/2}(\hat{\gamma} - \gamma_0) &= -\{\nabla_{\gamma\gamma} \bar{Q}_n(\hat{g}_{\tilde{\gamma}}, \tilde{\gamma})\}^{-1}\{n^{1/2}\nabla_\gamma \bar{Q}_n(\hat{g}_{\gamma_0}, \gamma_0)\} \\ &= -2\{\nabla_{\gamma\gamma} \bar{Q}_n(\hat{g}_{\tilde{\gamma}}, \tilde{\gamma})\}^{-1}\{\nabla_\gamma \bar{M}_n(\hat{g}_{\gamma_0}, \gamma_0)\}^{\mathrm{T}} \hat{W}_n\{n^{1/2}\bar{M}_n(\hat{g}_{\gamma_0}, \gamma_0)\}. \end{aligned} \tag{14}$$

For the asymptotic normality of $\hat{\gamma}$, we need the following additional conditions.

Condition 5. There is a vector of the functional $G(y, u, \delta, \omega)$ which is linear in $\omega = (\omega_1, \omega_2)^{\mathrm{T}}$ and such that:

(i) for small enough $\|\omega - \omega_0\|$, $\|\tilde{m}(y, u, \delta, \omega, \gamma_0) - \tilde{m}(y, u, \delta, \omega_0, \gamma_0) - G(y, u, \delta, \omega - \omega_0)\| \leqslant b(y, u, \delta)(\|\omega - \omega_0\|)^2$, where $\tilde{m}(y, u, \delta, \omega, \gamma)$ is the L-dimensional vector with lth component $\tilde{m}_l(y, u, \delta, \omega, \gamma) = I(z = l)[\delta\{1 + \exp(\gamma y)\omega_1(u)/\omega_2(u)\} - 1]$, $\omega_0 = [E(1 - \delta \mid u),\ E\{\delta \exp(\gamma_0 y) \mid u\}]^{\mathrm{T}}$, and $E\{b(y, u, \delta)\} < \infty$;
(ii) $\|G(y, u, \delta, \omega)\| \leqslant c(y, u, \delta)\|\omega\|$ and $E\{c(y, u, \delta)^2\} < \infty$;
(iii) there exists an almost everywhere continuous function $\upsilon(u)$ with $\int \|\upsilon(u)\| \,\mathrm{d}u < \infty$, $E\{G(y, u, \delta, \omega)\} = \int \upsilon(u)\omega(u) \,\mathrm{d}u$ for all $\|\omega\| \leqslant \infty$, and $E\{\sup_{\|\zeta\| \leqslant \epsilon} \|\upsilon(u + \zeta)\|^4\} < \infty$ for some $\epsilon > 0$.

Condition 6. For small enough $\|\omega - \omega_0\|$, $\tilde{m}(y, u, \delta, \omega, \gamma)$ is continuously differentiable in γ in a neighbourhood of γ_0, and there is $k(y, u, \delta)$ with $E\{k(y, u, \delta)\} < \infty$ such that $\|\nabla_\gamma \tilde{m}(y, u, \delta, \omega, \gamma) - \nabla_\gamma \tilde{m}(y, u, \delta, \omega_0, \gamma_0)\| \leqslant k(y, u, \delta)(\|\gamma - \gamma_0\|^\epsilon + \|\omega - \omega_0\|^\epsilon)$ for an $\epsilon > 0$, and $\Gamma = E\{\nabla_\gamma \tilde{m}(y, u, \delta, \omega_0, \gamma_0)\}$ exists and is of full rank.

THEOREM 2. *Assume that Conditions 1–6 hold. Then, as* $n \to \infty$, $n^{1/2}(\hat{\gamma} - \gamma_0) \to N(0, \Sigma)$ *in distribution, where* $\Sigma = (\Gamma^{\mathrm{T}} W \Gamma)^{-1}\Gamma^{\mathrm{T}} W \Omega W \Gamma (\Gamma^{\mathrm{T}} W \Gamma)^{-1}$, $\Omega = \mathrm{var}\{m(y, u, \delta, g_{\gamma_0}, \gamma_0) + \tau(y, u, \delta, \gamma_0)\}$, *and* $\tau(y, u, \delta, \gamma_0)$ *is a function given by* (S1) *in the Supplementary Material.*

THEOREM 3. *Suppose that Conditions 1–6 hold,* $f(u)$ *is bounded and the first and second derivatives of* $f(u)$ *are continuous and bounded, the probability function* $\pi(y, u, g_\gamma, \gamma)$ *satisfies* $\pi(y, u, g_\gamma, \gamma) > d_1 > 0$ *and* $E\{\pi(y, u, g_\gamma, \gamma) \mid u\} \neq 1$ *almost surely, and* $K(u) \leqslant d_2$ *for some* $d_2 > 0$ *on a closed interval centred at zero. Then, as* $n \to \infty$, $n^{1/2}(\hat{\mu} - \mu) \to N(0, V)$ *in distribution for some* $V > 0$.

The asymptotic variance V has a complicated form. Therefore, we recommend using the bootstrap to estimate V. Asymptotic results for other estimators using the inverse propensity weighting can be established similarly.

4. SIMULATIONS

4·1. *Comparison of estimators*

We conducted simulations to compare the finite-sample performance of the following six estimators of μ:

(i) $\hat{\mu}$, the proposed estimator in (12) using the estimated tilting parameter $\hat{\gamma}$ of (11);
(ii) $\hat{\mu}_{\gamma_0}$, the estimator in (12) using the true tilting parameter γ_0;

(iii) $\hat{\mu}_{\gamma_0-0\cdot3}$, the estimator in (12) using the wrong tilting parameter $\gamma_0 - 0\cdot3$;
(iv) the parametric estimator in Wang et al. (2014), $\tilde{\mu} = n^{-1}\sum_{i=1}^n \delta_i y_i\{1+\exp(\tilde{\alpha}+\tilde{\beta}^{\mathrm{T}}u_i+\tilde{\gamma}y_i)\}$, where a parametric model $\mathrm{pr}(\delta=1\mid y,u)=1/\{1+\exp(\alpha+\beta^{\mathrm{T}}u+\gamma y)\}$ is assumed and estimators $\tilde{\alpha}$, $\tilde{\beta}$ and $\tilde{\gamma}$ are obtained by the generalized method of moments;
(v) $\bar{y}_r$, the sample mean of the observed y_i;
(vi) $\bar{y}=\sum_{i=1}^n y_i/n$, the sample mean of y when there are no missing data, which is used as a benchmark.

The estimators $\hat{\mu}_{\gamma_0}$ and $\hat{\mu}_{\gamma_0-0\cdot3}$ are based on the approach adopted in Kim & Yu (2011), with $\hat{\mu}_{\gamma_0}$ using the correct tilting parameter and $\hat{\mu}_{\gamma_0-0\cdot3}$ using a wrong tilting parameter. All the simulated results are based on 1000 replications and sample sizes $n = 200$ and 500.

4·2. *Relative bias and standard deviation for one-dimensional u*

As in Wang et al. (2014), we considered a discrete instrument z having $L = 3$ categories. In the first simulation, $\mathrm{pr}(z_i=1)=0\cdot2$, $\mathrm{pr}(z_i=2)=0\cdot4$ and $\mathrm{pr}(z_i=3)=0\cdot4$. Given z_i, $u_i \sim N(z_i, 1)$. Given $z_i=1$ and u_i, $y_i = 1+0\cdot5(u_i-1)^2+\varepsilon_i$; given $z_i=2$ and u_i, $y_i=u_i^2+\varepsilon_i$; given $z_i=3$ and u_i, $y_i=2+(u_i-2)^2+\varepsilon_i$ where $\varepsilon_i\sim N(0,1)$. The second simulation is similar, except that z has two categories, with $\mathrm{pr}(z_i=1)=0\cdot4$ and $\mathrm{pr}(z_i=2)=0\cdot6$. The unconditional means of y in the two simulations are $\mu=3\cdot9$ and $\mu=3\cdot6$, respectively.

In addition to (u_i, y_i), we generated δ_i from Bernoulli distributions with probability π_i and considered the following six propensity models for π_i:

M_1: $\pi_i = 1/\{1+\exp(\alpha+\beta u_i)\}$, where $(\alpha,\beta)=(-0\cdot1,-0\cdot4)$ when $L=3$ and $(-0\cdot1,-0\cdot5)$ when $L=2$;
M_2: $\pi_i = 1/\{1+\exp(\alpha+\beta u_i+\gamma y_i)\}$, where $(\alpha,\beta,\gamma)=(0\cdot4,-0\cdot3,-0\cdot2)$ when $L=3$ and $(0\cdot1,-0\cdot2,-0\cdot2)$ when $L=2$;
M_3: $\pi_i = 1/[1+\exp\{\alpha+\beta\sin(u_i)+\gamma y_i\}]$, where $(\alpha,\beta,\gamma)=(0\cdot1,-0\cdot1,-0\cdot3)$ when $L=3$ and $(0\cdot1,-0\cdot1,-0\cdot2)$ when $L=2$;
M_4: $\pi_i = 1/\{1+\exp(\alpha+\beta u_i^2+\gamma y_i)\}$, where $(\alpha,\beta,\gamma)=(0\cdot5,-0\cdot2,-0\cdot1)$ when $L=3$ and $(0\cdot2,-0\cdot3,-0\cdot1)$ when $L=2$;
M_5: $\pi_i = 1/\{1+\exp(\alpha+\beta_1 u_i^2+\beta_2 u_i^{-2}+\gamma y_i)\}$, where $(\alpha,\beta_1,\beta_2,\gamma)=(0\cdot5,-0\cdot2,-0\cdot1,-0\cdot05)$ when $L=3$ and $(0\cdot5,-0\cdot3,-0\cdot1,-0\cdot1)$ when $L=2$;
M_6: $\pi_i = 1/[1+\exp\{\alpha+\beta\exp(u_i)+\gamma y_i\}]$, where $(\alpha,\beta,\gamma)=(0\cdot5,-0\cdot1,-0\cdot1)$ when $L=3$ and $(0\cdot2,-0\cdot15,-0\cdot1)$ when $L=2$.

Of these scenarios, M_1 is an ignorable missing data case, and M_2–M_6 represent five different kinds of nonignorable missing data cases. In particular, M_1 and M_2 satisfy the response probability assumption in Wang et al. (2014). The parameter values under the missing data models were chosen so that the unconditional rate of missing data was about 30%.

As in Chen et al. (2015), the nonparametric kernel regression estimator of (9) was computed using a Gaussian kernel $K(u)=\exp(-u^2/2)/(2\pi)^{1/2}$. The bandwidths were selected depending on l, i.e., $h_l = 1\cdot5\,\hat{\sigma}_{u_l} n_l^{-1/3}$, where $\hat{\sigma}_{u_l}$ and n_l are the estimated standard deviation and the sample size of u_i based on the lth category sample.

Table 1 reports the simulated relative biases and standard deviations of the six point estimators for $n = 200$. It can be seen that $\bar{y}_r$ is biased because of the fact that missingness is not completely random. The proposed estimator $\hat{\mu}$ and $\hat{\mu}_{\gamma_0}$ with the correct parameter value γ_0 have negligible biases in all cases, and their standard deviations are comparable to and larger than those of $\bar{y}$. The estimator $\hat{\mu}_{\gamma_0-0\cdot3}$ uses a wrong value of γ_0 and thus is biased, although it is less biased

Table 1. *Relative biases and standard deviations of six estimators of μ; all values have been multiplied by* 100

				Estimator					
Dim(u)	L	Model	Quantity	$\hat{\mu}$	$\hat{\mu}_{\gamma_0}$	$\hat{\mu}_{\gamma_0-0\cdot3}$	$\tilde{\mu}$	$\bar{y}_r$	$\bar{y}$
1	3	M_1	RB	−0·8	−0·8	−4·6	−0·5	7·9	0·1
			SD	25·5	25·5	25·3	39·9	29·8	24·3
		M_2	RB	0·3	−0·4	−3·5	1·1	17·2	−0·2
			SD	26·0	25·9	26·6	45·8	32·1	25·0
		M_3	RB	1·5	−0·2	−3·7	0·7	16·6	0·2
			SD	25·8	25·9	26·6	37·4	32·3	25·3
		M_4	RB	−0·7	−0·1	−2·9	7·3	18·9	−0·1
			SD	26·1	25·9	26·3	55·2	32·2	25·0
		M_5	RB	−0·5	0·3	−3·2	7·6	14·6	0·4
			SD	26·2	26·2	26·0	46·6	32·7	25·4
		M_6	RB	−1·0	−0·2	−2·7	5·8	18·6	−0·1
			SD	26·9	26·8	26·1	55·3	33·2	26·0
1	2	M_1	RB	−1·0	−0·8	−5·0	1·4	11·6	−0·1
			SD	28·6	28·4	28·2	54·9	35·4	27·6
		M_2	RB	0·6	−0·4	−4·3	3·6	20·4	−0·1
			SD	28·4	28·6	28·8	62·0	37·8	28·3
		M_3	RB	0·8	−0·9	−5·9	3·7	20·5	−0·2
			SD	28·5	28·5	28·5	62·8	38·1	27·6
		M_4	RB	−0·7	0·2	−2·5	9·2	21·5	−0·2
			SD	27·6	27·6	28·0	63·1	35·4	27·2
		M_5	RB	−1·2	−0·5	−3·3	7·2	18·3	−0·7
			SD	27·4	27·3	27·7	53·1	35·1	26·7
		M_6	RB	−0·3	0·4	−2·3	6·3	20·8	0·2
			SD	28·4	28·1	28·3	58·2	35·6	27·6
2	3	M_1	RB	−3·0	−3·0	−6·1	−0·1	5·6	0·5
			SD	28·3	28·3	27·8	39·3	32·9	27·1
		M_2	RB	−1·2	−2·0	−4·7	0·8	12·5	0·1
			SD	27·9	28·1	28·2	41·0	33·3	26·9
		M_3	RB	−2·3	−2·2	−4·4	7·0	16·2	−0·1
			SD	29·1	29·2	29·4	60·5	34·3	27·7
		M_4	RB	−1·9	−1·9	−4·5	3·2	15·0	−0·1
			SD	27·0	27·2	27·3	52·6	35·1	27·9

RB, relative bias; SD, standard deviation.

than $\bar{y}_r$. The estimator $\tilde{\mu}$ is nearly unbiased under the missing data models M_1 and M_2 when the parametric model on $\text{pr}(\delta = 1 \mid y, u)$ is correct, but it is biased under M_3–M_6 when the parametric model on $\text{pr}(\delta = 1 \mid y, u)$ is wrong. Moreover, the standard deviation of $\tilde{\mu}$ is larger than that of $\hat{\mu}$ even when the parametric model on $\text{pr}(\delta = 1 \mid y, u)$ is correct, which could be due to the fact that the estimating equation in Wang et al. (2014) is not optimal. Similar results for $n = 500$ are given in the Supplementary Material.

4·3. *Relative bias and standard deviation for two-dimensional u*

We also considered the case of a two-dimensional $u_i = (u_{i1}, u_{i2})^{\mathrm{T}}$ and a z having $L = 3$ categories, with $\text{pr}(z_i = 1) = 0{\cdot}2$, $\text{pr}(z_i = 2) = 0{\cdot}4$ and $\text{pr}(z_i = 3) = 0{\cdot}4$. Given z_i, u_{i1} and u_{i2} are independently generated from normal distributions $N(z_i, 1)$ and uniform distributions $\text{Un}(0, z_i)$. Given $z_i = 1$, u_{i1} and u_{i2}, $y_i = 1 + 0{\cdot}5(u_{i1} - 1)^2 + 0{\cdot}5(u_{i2} - 1)^2 + \varepsilon_i$; given $z_i = 2$, u_{i1} and u_{i2}, $y_i = u_{i1}^2 + u_{i2}^2 + \varepsilon_i$; given $z_i = 3$, u_{i1} and u_{i2}, $y_i = 2 + (u_{i1} - 2)^2 + (u_{i2} - 2)^2 + \varepsilon_i$ where

Table 2. *Coverage probabilities and standard errors for $n = 200$; all values have been multiplied by 100*

				Estimator					
Dim(u)	L	Model	Quantity	$\hat{\mu}$	$\hat{\mu}_{\gamma_0}$	$\hat{\mu}_{\gamma_0-0\cdot3}$	$\tilde{\mu}$	$\bar{y}_r$	$\bar{y}$
1	3	M_1	CP	95·0	92·2	84·5	93·1	85·9	94·9
			SE	25·8	25·2	24·9	39·1	29·8	24·4
		M_2	CP	94·6	92·2	83·0	93·3	43·9	95·1
			SE	25·3	25·3	24·9	41·6	31·3	24·5
		M_3	CP	93·2	94·3	89·7	94·9	44·1	93·7
			SE	25·3	25·3	25·5	37·1	31·0	24·5
1	2	M_1	CP	94·1	93·7	88·0	95·5	79·5	93·9
			SE	28·2	27·8	27·6	55·1	34·6	27·1
		M_2	CP	94·2	95·3	89·9	94·1	44·3	93·8
			SE	27·9	27·7	27·8	63·0	35·5	26·9
		M_3	CP	93·7	95·8	86·8	94·1	48·2	94·4
			SE	27·9	27·8	27·8	69·1	37·2	26·9
2	3	M_1	CP	92·1	88·9	75·8	94·8	88·8	93·7
			SE	28·3	27·9	27·7	41·7	31·9	26·7
		M_2	CP	95·6	91·4	84·9	93·3	55·0	92·9
			SE	28·4	27·9	27·9	40·0	32·7	26·7
		M_3	CP	93·7	92·0	86·9	86·2	29·7	93·7
			SE	28·6	28·3	28·4	53·7	33·8	27·0

CP, coverage probability; SE, standard error.

$\varepsilon_i \sim N(0, 1)$. The unconditional mean is $\mu = 4{\cdot}9$. We considered the following four models for π_i:

M_1: $\pi_i = 1/\{1 + \exp(\alpha + \beta_1 u_{i1} + \beta_2 u_{i2})\}$, where $(\alpha, \beta_1, \beta_2) = (0{\cdot}1, -0{\cdot}3, -0{\cdot}3)$;
M_2: $\pi_i = 1/\{1 + \exp(\alpha + \beta_1 u_{i1} + \beta_2 u_{i2} + \gamma y_i)\}$, where $(\alpha, \beta_1, \beta_2, \gamma) = (0{\cdot}1, -0{\cdot}2, -0{\cdot}2, -0{\cdot}1)$;
M_3: $\pi_i = 1/[1 + \exp(\alpha + \beta_1 u_{i1}^2 + \beta_2 u_{i2}^2 + \gamma y_i)]$, where $(\alpha, \beta_1, \beta_2, \gamma) = (0{\cdot}8, -0{\cdot}2, -0{\cdot}2, -0{\cdot}1)$;
M_4: $\pi_i = 1/[1 + \exp\{\alpha + \beta_1 \exp(u_{i1}) + \beta_2 \exp(u_{i2}) + \gamma y_i\}]$, where $(\alpha, \beta_1, \beta_2, \gamma) = (0{\cdot}2, -0{\cdot}05, -0{\cdot}05, -0{\cdot}05)$.

As in Tang et al. (2014), we chose the kernel function to be the product kernel, $K(u_{i1}, u_{i2}) = K(u_{i1})K(u_{i2})$, and chose the bandwidth to be $h_l = 1{\cdot}5\,\hat{\sigma}_{u_{1l}} n_l^{-1/3}$ where $\hat{\sigma}_{u_{1l}}$ is the standard deviation of observations of u_{i1} in the lth category sample. The relative biases and standard deviations of the six point estimators of μ obtained from this simulation are shown in Table 1. Conclusions about and comparisons of the estimators' performance are similar to those in the case of one-dimensional u.

4·4. *Coverage probability and standard error*

Under models M_1–M_3, we further examined the confidence intervals of μ based on the six estimators, the normal approximation, and the bootstrap estimators of the asymptotic variances. For $n = 200$, Table 2 shows the simulation results on the coverage probabilities of confidence intervals and the standard errors, i.e., the square roots of bootstrap variance estimators based on 50 bootstrap replications.

A few conclusions can be drawn from Table 2. First, the coverage probability based on the proposed $\hat{\mu}$ is close to the nominal level 0·95, and is quite comparable to the method based on $\bar{y}$

Table 3. *Estimates (with standard errors in parentheses) of monthly income in* 2006 *from the Korean Labor and Income Panel Study data*

Instrument z	L	$\tilde{\mu}$	$\hat{\mu}$
age, education, gender	12	183·9 (2·6)	186·4 (2·3)
age, education	6	185·4 (3·0)	186·3 (2·4)
age, gender	6	183·1 (3·8)	186·3 (2·4)
education, gender	4	183·2 (4·5)	183·6 (2·3)
age	3	196·2 (6·3)	186·1 (2·2)
education	2	188·2 (5·9)	185·6 (2·4)
gender	2	186·1 (5·5)	185·7 (2·2)

assuming no missing data. The main price paid for missing data is the increased standard error, so that the confidence intervals are wider due to missing data. Second, in terms of the coverage probability, the method based on $\hat{\mu}_{\gamma_0}$ with the true value γ_0 surprisingly did not perform well in some cases, but the reason for this is not clear. Third, when a wrong value $\gamma_0 - 0{\cdot}3$ is used, the method based on $\hat{\mu}_{\gamma_0-0.3}$ does not have good coverage probability, although it is better than the coverage probability obtained from the naive method of using the observed sample mean $\bar{y}_r$. This shows that the bias may cause some problems, although the relative bias seems low. Finally, the method based on $\tilde{\mu}$ of Wang et al. (2014) performs well under models M_1 and M_2 where the parametric model for propensity is correct, but may not perform well under M_3 when the propensity model is wrong.

5. Example

We applied the proposed method to the Korean Labor and Income Panel Study dataset (Wang et al., 2014), a detailed description of which can be found at www.kli.re.kr/klips/en/about/introduce.jsp. The dataset includes $n = 2506$ regular wage earners. The variable of interest y is the monthly income in 2006, with covariates gender, age group, education level, and monthly income in 2005. The variable y has approximately 35% missing values, while all covariate values are observed. Income in 2005 was considered a continuous covariate and treated as u or part of u. The discrete covariates are defined as follows: gender has two categories, male and female; age group has three categories, age < 35, $35 \leqslant$ age < 51, and age $\geqslant 51$; education level has two categories, up to high school and beyond high school. We think it reasonable to assume that, given 2005 income and 2006 income, the missing data propensity does not depend on some or all of gender, age group and education level. But it is not clear which subset of gender, age group and education level is the best choice of the instrument z. To investigate the effect of using different instruments, we computed the proposed estimates $\hat{\mu}$ with all seven possible instrument subsets. We also computed $\tilde{\mu}$ of Wang et al. (2014).

The estimates and their standard errors based on the bootstrap are reported in Table 3 for different choices of z. Our proposed estimator is insensitive to these choices. Compared with $\tilde{\mu}$, our proposed method has smaller standard errors, consistent with the simulation results in § 4. The variation in $\tilde{\mu}$ values with different choices of the instrument is also larger.

6. Discussion

To apply the proposed method, we need to choose an instrument z to meet two conditions: (i) z must be related to the study variable y, i.e., $f(y \mid x)$ depends on z; (ii) z can be excluded from the propensity, i.e., (2) holds. In some applications, categorical covariates such as age group, gender, race and education level are related to the study variable y but almost unrelated to the propensity

when y and other covariates are conditioned; see § 5. Let us discuss how to choose an instrument satisfying (i) and (ii). As argued in § 2·1, if z does not satisfy (i), then the equations constructed based on (7) and (8) reduce to a single equation, so the generalized method of moments algorithm does not converge, regardless of whether (ii) is satisfied. Hence, a z not satisfying (i) can easily be excluded from the search for an instrument. Next, if (7) and (8) are based on a z satisfying (i) but not (ii), then the resulting generalized moment estimator of $\pi(y, x)$, denoted by $\tilde{\pi}(y, u)$, can be computed but does not converge to the true propensity $\pi(y, x)$. Consequently,

$$D = \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i x_i}{\tilde{\pi}(y_i, u_i)} - \frac{1}{n} \sum_{i=1}^{n} x_i \right\| \tag{15}$$

does not converge to zero in probability. On the other hand, if z satisfies both (i) and (ii), then the quantity in (15) converges to zero in probability. Therefore, we can choose a z that satisfies (i) and minimizes D in (15). Some simulation results on selecting an instrument using this idea are given in the Supplementary Material.

Note that (15) focuses on the estimation bias, so it may not tell us which subset of instruments is best if there are multiple instruments. Therefore we need another criterion, probably related to the efficiency of estimators, for choosing the best subset of instruments; this requires further research. Another problem that calls for more research is how to determine an appropriate parametric model for the effect of y in the propensity. Some simulation results on the performance of the proposed estimator $\hat{\mu}$ when a working model is $\pi(y, x) = 1/[1 + \exp\{g(u) + \gamma y\}]$ but the true model is $\pi(y, x) = 1/[1 + \exp\{g(u) + \gamma y^2\}]$ are given in the Supplementary Material.

Our method requires the estimation of the quantity in (13), and we propose using the kernel estimator given by (9). Any other feasible nonparametric method, such as the generalized additive model (Xie et al., 2011) or the sufficient dimension reduction technique (Li, 1991; Cook & Weisberg, 1991; Cook & Li, 2002), can be applied to obtain an estimator of $\exp\{g_\gamma(u)\}$ for fixed γ, especially when the dimension of u is not low.

ACKNOWLEDGEMENT

We are grateful to the editor, associate editor and two referees for comments that led to significant improvements of the paper. The authors' research was partially supported by the Chinese 111 Project, Shanghai Scientific Grant, the U.S. National Science Foundation, the National Natural Science Foundation and the Postdoctoral Science Foundation of China. L. Wang is also affiliated with the School of Statistics, East China Normal University.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online contains some proofs and further simulation results.

REFERENCES

CHEN, X., WAN, A. T. & ZHOU, Y. (2015). Efficient quantile regression analysis with missing observations. *J. Am. Statist. Assoc.* **110**, 723–41.

COOK, R. D. & LI, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.* **30**, 455–74.

COOK, R. D. & WEISBERG, S. (1991). Discussion of "Sliced inverse regression for dimension reduction". *J. Am. Statist. Assoc.* **86**, 328–32.

FITZMAURICE, G. M., MOLENBERGHS, G. & LIPSITZ, S. R. (1995). Regression models for longitudinal binary responses with informative drop-outs. *J. R. Statist. Soc.* B **57**, 691–704.

GREBLICKI, W., KRZYZAK, A. & PAWLAK, M. (1984). Distribution-free pointwise consistency of kernel regression estimate. *Ann. Statist.* **12**, 1570–75.

HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029–54.

KHAN, S. & POWELL, J. L. (2001). Two-step estimation of semiparametric censored regression models. *J. Economet.* **103**, 73–110.

KIM, J. K. & YU, C. L. (2011). A semiparametric estimation of mean functionals with nonignorable missing data. *J. Am. Statist. Assoc.* **106**, 157–65.

LI, K. C. (1991). Sliced inverse regression for dimension reduction. *J. Am. Statist. Assoc.* **86**, 316–27.

LITTLE, R. J. A. (1985). A note about models for selectivity bias. *Econometrica* **53**, 1469–74.

MOLENBERGHS, G. & KENWARD, M. G. (2007). *Missing Data in Clinical Studies*. New York: Wiley.

NEWEY, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Economet. Theory* **10**, 233–53.

NEWEY, W. K. & MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, vol. 4, R. F. Engle & D. L. McFadden, eds. Amsterdam: Elsevier, pp. 2111–245.

QIN, J., LEUNG, D. & SHAO, J. (2002). Estimation with survey data under nonignorable nonresponse or informative sampling. *J. Am. Statist. Assoc.* **97**, 193–200.

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

TANG, G., LITTLE, R. J. A. & RAGHUNATHAN, T. E. (2003). Analysis of multivariate missing data with nonignorable nonresponse. *Biometrika* **90**, 747–64.

TANG, N., ZHAO, P. & ZHU, H. (2014). Empirical likelihood for estimating equations with nonignorably missing data. *Statist. Sinica* **24**, 723–47.

WANG, S., SHAO, J. & KIM, J. K. (2014). An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statist. Sinica* **24**, 1097–116.

XIE, H., QIAN, Y. & QU, L. (2011). Semiparametric approach for analyzing nonignorable missing data. *Statist. Sinica* **21**, 1881–99.

YANG, F., LORCH, S. A. & SMALL, D. S. (2014). Estimation of causal effects using instrumental variables with nonignorable missing covariates: Application to effect of type of delivery NICU on premature infants. *Ann. Appl. Statist.* **8**, 48–73.

ZHAO, H., ZHAO, P. & TANG, N. (2013). Empirical likelihood inference for mean functionals with nonignorably missing response data. *Comp. Statist. Data Anal.* **66**, 101–16.

[*Received April* 2015. *Revised November* 2015]