Identification and Doubly Robust Estimation of Nonignorable Missing Data With an Ancillary Variable

Wang Miao

Beijing University

mwfy@pku.edu.cn

November 13, 2015

Table of contents

Introduction

Y outcome, R missingness indicator, X a vector of covariates.

- The interest is the mean of an outcome variable missing not at random (MNAR)

$$Y \not\perp\!\!\!\perp R|X.$$

Figure : MNAR Versus MAR

- MNAR is common in empirical studies.
  Examples: sensitivity questions in surveys.

Introduction
MNAR lacks identification.

Identification means that the observed data likelihood uniquely determines the joint distribution.

The observed data likelihood is $P(y, r = 1|x)^r P(r = 0|x)^{1-r}$. The unknown parameter is $P(y, r|x)$.

- Nonparametric models are not identifiable in general.
- Even some parametric models are not identifiable.

Example

Normal-logit models (Miao et al., 2014; Wang et al., 2014).

$$Y|x \sim N(\beta_0 + \beta_1 x, \sigma^2), \quad \text{logit } P(r = 1|y, x) = \alpha_0 + \alpha_1 y + \alpha_2 x.$$

Introduction
Mitigation strategies

- Stringently parametric models (the Heckman Selection Model);
- the instrumental variable approach $W \perp\!\!\!\perp Y|X$, $W \not\perp\!\!\!\perp R|X$;

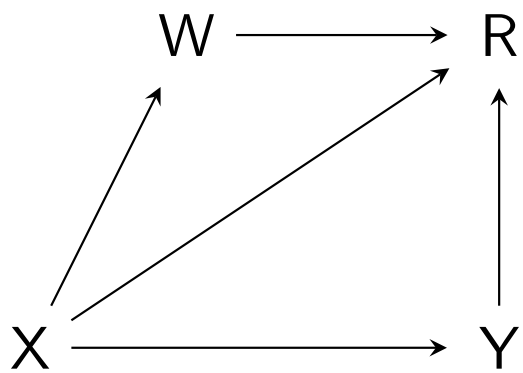


Figure : The instrumental variable.

Recently developed approach

- The ancillary variable approach $Z \perp\!\!\!\perp R|(Y, X)$, $Z \not\perp\!\!\!\perp Y|X$. (Wang et al., 2014; Zhao & Shao, 2014; Miao et al., 2015)

Introduction
Definition of ancillary variable

- We call Z an ancillary variable if Z satisfies

$$Z \perp\!\!\!\perp R|(Y,X), \quad Z \not\perp\!\!\!\perp Y|X.$$

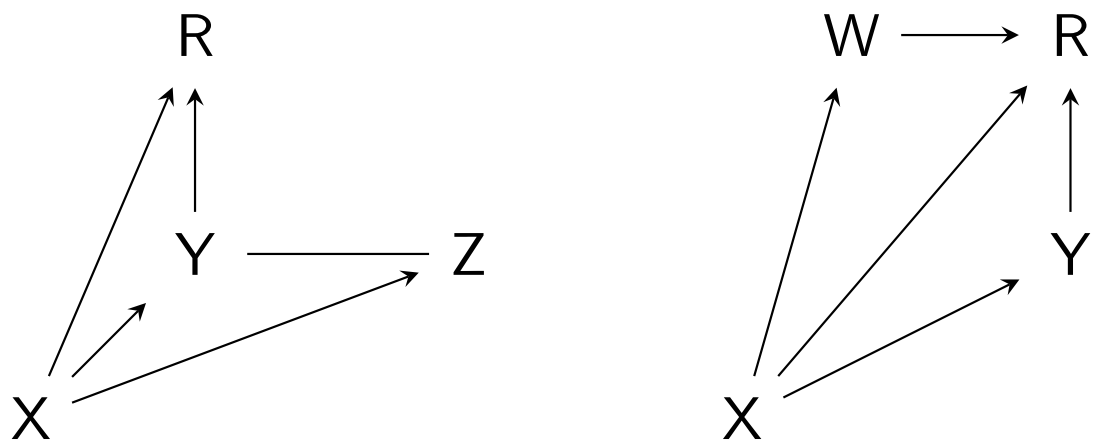


Figure : Ancillary variable Versus instrumental variable.

- Examples: proxy, mismeaured version of the outcome. Student mental health example from Ibrahim et al. (2001): Y teacher's assessment, Z parental assessment.

Identification
Use of an ancillary variable

With an ancillary variable,

- the observed data likelihood $P(z, y, r = 1|x)^r P(z, r = 0|x)^{1-r}$;
- the unknown parameter $P(z, y|x)$, $P(r|y, x)$;
- the observed data likelihood is a functional of the parameter

$$P(z, y, r = 1|x) = P(z, y|x)P(r = 1|y, x),$$
$$P(z, r = 0, x) = \int_y P(z, y|x)P(r = 1|y, x)dy.$$

An ancillary variable introduces more restrictions than unknown parameters.

Identification
Binary outcome case

Table : Binary outcome case

	With a binary Z	Without Z
Para	$P(z, y)$, $P(r\|y)$	$P(y, r)$
#	5	3
Restrictions	$P(z, y, r = 1)$, $P(z, r = 0)$	$P(y, r = 1)$, $P(r = 0)$
#	5	2

The ancillary variable assumption is enough for a binary outcome (Ma et al., 2003; Miao et al., 2015).

Identification
Continuous outcome case

Continuous outcome case needs a continuous ancillary variable and further model restrictions.

Semiparametric models (Miao et al., 2015)

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

with unrestricted functions μ_r and σ_r, and density functions P_r.

Theorem

Suppose either $P(y|z,x,r=1)$ *or* $P(y|z,x,r=0)$ *follows the above model with the corresponding density function* P_r *satisfying the regularity conditions. Then the joint distribution of* $(Z,Y,R|X)$ *is identified.*

Identification
Lack of identification is not an issue.

With an ancillary variable

- for a continuous outcome, the model class includes many commonly-used models.

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

Example

Gaussian models, Student-t models.

- a binary outcome case is always identifiable;

Estimation
Parametrization (Chen, 2007)

$$
\begin{aligned}
P(z, y, r|x) &= c(x) \exp\{(1-r) OR(y|x)\} P(r|y=0, x) P(z, y|r=1, x) \\
OR(y|x) &= \log \frac{P(r=0|y, x) P(r=1|y=0, x)}{P(r=0|y=0, x) P(r=1|y, x)}.
\end{aligned}
$$

The baseline propensity score $P(r=1|y=0, x)$; the baseline outcome density $P(z, y|r=1, x)$; the log odds ratio function encodes the degree of the missingness process departs from MAR.

Example

For a logistic missingness model
logit $P(r=1|y, x) = \alpha_0 + \alpha_1 y + \alpha_2 x$, $OR(y|x) = -\alpha_1 y$.

Estimation
Parametrization (Chen, 2007)

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

- such parametrization extends that under MAR.
  $OR(y|x)=0$, $P(r|y=0,x)=P(r|x)$.
- We specify separate parametric models $OR(y|x;\gamma)$,
  $P(r=1|y=0,x;\alpha)$, $P(z,y|r=1,x;\beta)$.

Estimation
Estiation of nuisance parameters

- We specify separate parametric models $OR(y|x;\gamma)$, $P(r=1|y=0,x;\alpha)$, $P(z,y|r=1,x;\beta)$.
- We estimate (α,β,γ) by solving

$$\mathbb{P}_n\{\partial \log\{P(z,y|r=1,x;\widehat{\beta})\}/\partial\beta\}=0,$$

$$\mathbb{P}_n\left[\{W(x,y;\widehat{\alpha},\widehat{\gamma})r-1\}\left\{\begin{array}{c} z-\mathbb{E}(Z|r=0,x;\widehat{\beta},\widehat{\gamma}) \\ x \end{array}\right\}\right]=0,$$

$\mathbb{P}_n$ stands for empirical mean; $W(x,y,\widehat{\alpha},\widehat{\gamma})$ is the inverse probability weight function.

Estimation
Estiation of nuisance parameters

- Perfect version

$$\mathbb{P}_n \left[\{ W(x, y; \widehat{\alpha}, \widehat{\gamma}) r - 1 \} \left\{ \begin{array}{c} y \\ x \end{array} \right\} \right] = 0,$$

- Z takes the place of Y.

$$\mathbb{P}_n \left[\{ W(x, y; \widehat{\alpha}, \widehat{\gamma}) r - 1 \} \left\{ \begin{array}{c} z - \mathbb{E}(Z | r = 0, x; \widehat{\beta}, \widehat{\gamma}) \\ x \end{array} \right\} \right] = 0,$$

$\mathbb{P}_n$ stands for empirical mean; $W(x, y, \widehat{\alpha}, \widehat{\gamma})$ is the inverse probability weight function.

Estimation
Non-doubly robust estimators

- IPW estimator and Horvitz-Thompson estimator

$$\begin{aligned}\widehat{\mu}_{ipw} &= \mathbb{P}_n\left\{W\left(x,y;\widehat{\alpha},\widehat{\gamma}\right)ry\right\},\\ \widehat{\mu}_{HT} &= \mathbb{P}_n\left\{\frac{W(x,y;\widehat{\alpha},\widehat{\gamma})r}{\mathbb{P}_n\{W(x,y;\widehat{\alpha},\widehat{\gamma})r\}}y\right\}.\end{aligned}$$

- Regression based estimator

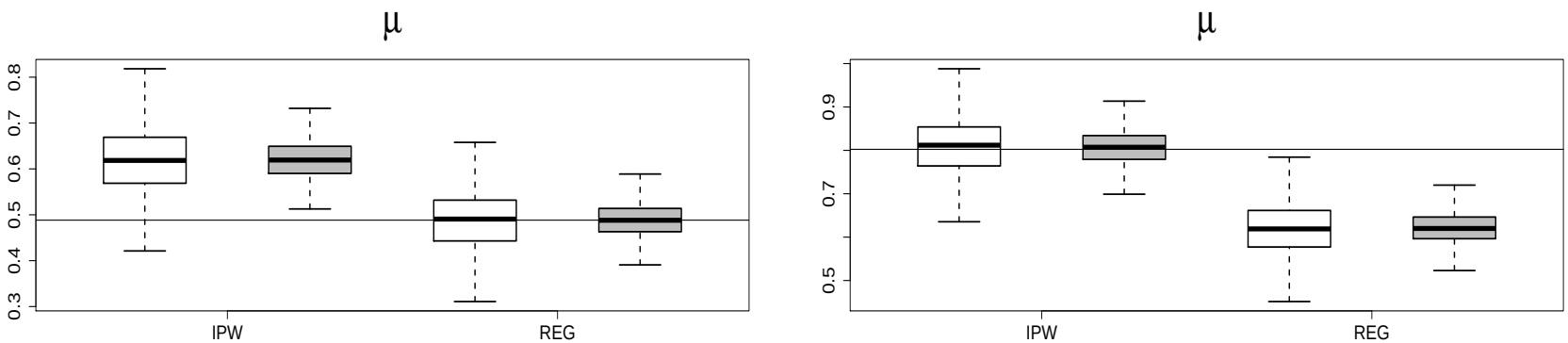
$$\widehat{\mu}_{reg} = \mathbb{P}_n\{(1-r)M_0(x;\widehat{\beta},\widehat{\gamma}) + ry\}.$$

$M_0(x;\widehat{\beta},\widehat{\gamma})$ is the estimated conditional outcome mean of the missing part, i.e. $\mathbb{E}(Y|r=0,x;\widehat{\beta},\widehat{\gamma})$.

Estimation

Non-doubly robust estimators

We assume that the log odds ratio model is correct. Non-doubly robust estimators are biased if the required baseline model is incorrect.



(a) Correct baseline outcome model (b) Correct baseline propensity score

Figure : Boxplots of the estimators for the mean of a normal outcome.

Note: in each boxplot, white boxes are for sample size 500, and gray ones for 1500. The horizontal line marks the true value of the parameter.

Estimation
Doubly robust estimators

- Doubly robust estimators combine IPW and Regression based estimation.
- Double robustness: assuming a correct $OR(y|x;\gamma)$, a doubly robust estimator is consistent if either of the baseline model is correct;
- a doubly robust estimator is biased if both baseline models are incorrect.

Doubly robust estimators
Horvitz-Thompson estimator with extended weights

- Extended baseline propensity score model and weight function

$$P_{ext}(r = 1|y = 0, x; \phi) = P(r = 1|y = 0, x; \widehat{\alpha}) \text{ only at } \phi = 0,$$

$$W_{ext}(x, y; \widehat{\alpha}, \widehat{\gamma}, \phi) = 1 + \exp\{OR(y|x; \widehat{\gamma})\} \frac{P_{ext}(r = 0|y = 0, x; \phi)}{P_{ext}(r = 1|y = 0, x; \phi)}.$$

- Estimation of the extended weight function

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

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

Doubly robust estimators
Horvitz-Thompson estimator with extended weights

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

Versus

$$\widehat{\mu}_{HT} = \mathbb{P}_n\left\{ \frac{W(x, y; \widehat{\alpha}, \widehat{\gamma})r}{\mathbb{P}_n\{W(x, y; \widehat{\alpha}, \widehat{\gamma})r\}} y \right\}.$$

Doubly robust estimators
Regression estimator with an extended outcome model

- Extended outcome model
  $M_{0,ext}(x; \widehat{\beta}, \widehat{\gamma}, \psi) = M_0(x; \widehat{\beta}, \widehat{\gamma})$ only at $\psi = 0$.
- Estimation of ψ

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

Doubly robust estimators
Regression estimator with an extended outcome model

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

Versus

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

Doubly robust estimators
Regression estimator with residual bias correction

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

- under MAR, $OR(y|x) = 0$, $P(r = 1|x, y = 0) = P(r = 1|x)$, and $M_0(x) = \mathbb{E}(Y|x, r = 0) = \mathbb{E}(Y|x) = M(x)$;

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

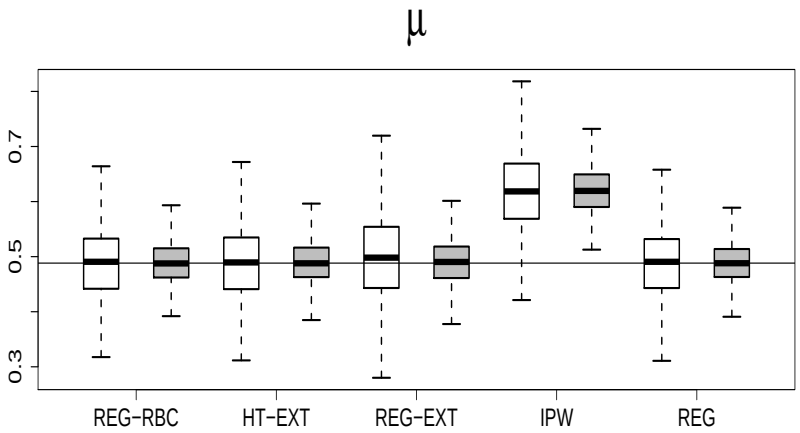
Doubly robust estimators
Double robustness

Theorem

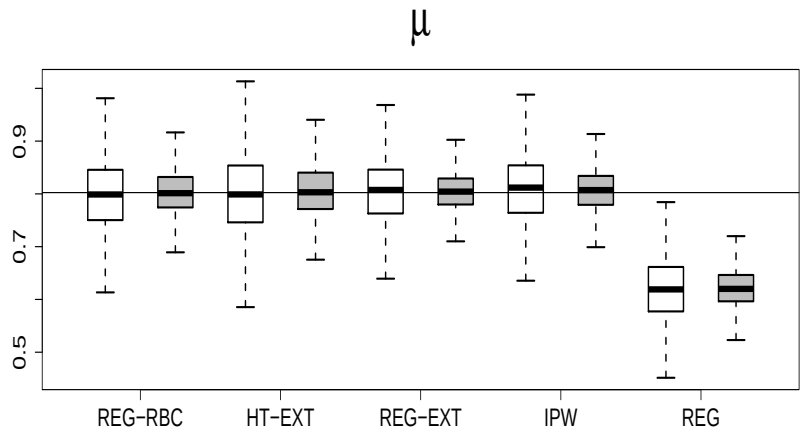
Suppose that the log odds ratio model $OR(y|x;\gamma)$ *is correctly specified, then the estimators* $\widehat{\mu}_{REG-RBC}$, $\widehat{\mu}_{HT-EXT}$ *and* $\widehat{\mu}_{REG-EXT}$ *are consistent if either* $P(z,y|r=1,x;\beta)$ *or* $P(r=1|y=0,x;\alpha)$ *is correctly specified.*

- By use of the log odds ratio model $OR(y|x;\gamma)$ we relax previous stringent restrictions, for example, a null value of zero under MAR.

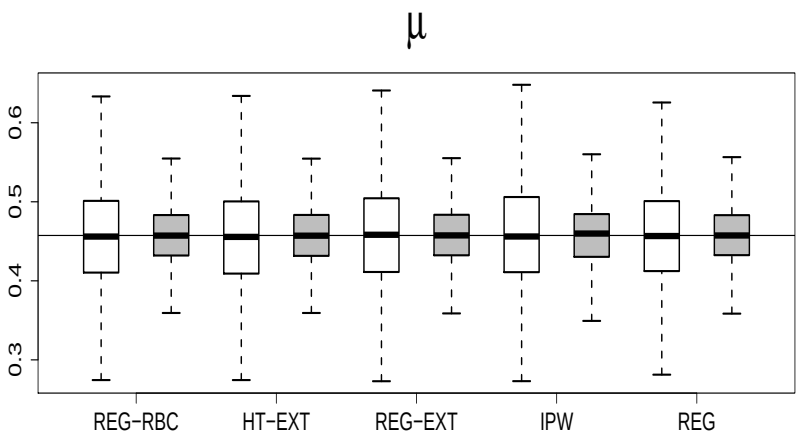
Doubly robust estimators
Double robustness



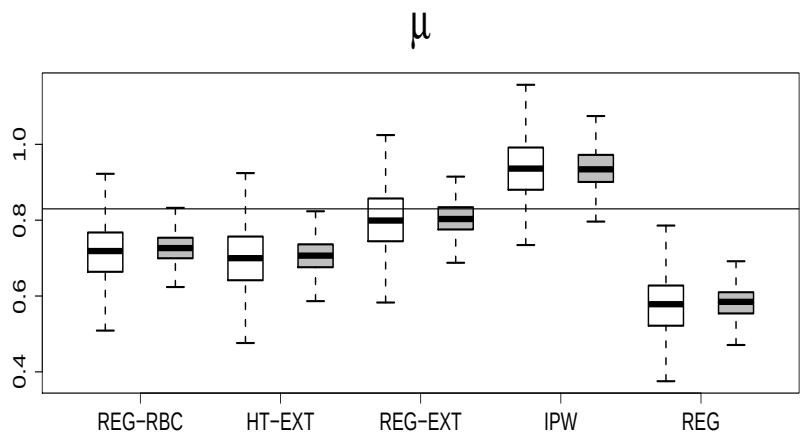
(a) FT



(b) TF



(c) TT



(d) FF

A comparison of the estimators
Boundedness

Boundedness (Robins et al., 2007; Tan, 2010): an estimator falls in the parameter space for the outcome mean almost surely.

- Bounded estimators are preferred when the inverse probability weights are highly variable.

Boundedness	$\widehat{\mu}_{REG-RBC}$	$\widehat{\mu}_{HT-EXT}$	$\widehat{\mu}_{REG-EXT}$
	No	Yes	it depends

- In general settings, no recommendation can be made.

A comparison of the estimators
Check model specification

- Extended outcome model
  $M_{0,ext}(x;\psi) = M_0(x;\widehat{\beta},\widehat{\gamma})$ only at $\psi = 0$.
  Extended baseline propensity score model
  $P_{ext}(r = 1|y = 0, x;\phi) = P(r = 1|y = 0, x;\widehat{\alpha})$ only at $\phi = 0$.
- If a baseline model is correct, the corresponding extended model must have $\phi = 0$ or $\psi = 0$.
- Test $\phi = 0$ or $\psi = 0$ to check if the baseline model is correct.

A comparison of the estimators
Check model specification

Table : Empirical type I error and power

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

Note: The baseline propensity score model is correct only in (b) and (c); the baseline outcome model is correct only in (a) and (c). The result of each situation includes two rows, of which the first row stands for sample size 500, and the second for 1500.

Conclusion

- We use an ancillary variable to help identify the models. Lack of identification is not an issue in many situations.
- We propose three doubly robust estimators of the outcome mean. They combine the baseline models to achieve double robustness.
- The doubly robust estimators can be used to test if the baseline models are correctly specified.

Acknowledgment

China Scholarship Council, and US National Institute of Health.

Eric Tchetgen Tchetgen

Zhi Geng

Jae-kwang Kim

References

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

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

Ma, W. Q., Geng, Z. & Hu, Y. H. (2003). Identification of graphical models for nonignorable nonresponse of binary outcomes in longitudinal studies. *Journal of multivariate analysis* **87**, 24–45.

Miao, W., Ding, P. & Geng, Z. (2014). Identifiability of normal and normal mixture models with nonignorable missing data **Accepted by JASA**.

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

Robins, J., Sued, M., Lei-Gomez, Q. & Rotnitzky, A. (2007). Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable. *Statistical Science* , 544–559.

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

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

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