Online Supplement to "Identification and Inference With Nonignorable Missing Covariate Data"

Wang Miao and Eric Tchetgen Tchetgen

Peking University and Harvard University

July 27, 2017

This supplement includes identification results for the pattern-mixture parametrization, efficiency issue for (8), useful lemmas, and proofs of the theorems.

A. Identification Results for the Pattern-Mixture parametrization

Considering a model $\mathrm{pr}(x, y, z, r; \theta)$ indexed by θ, we assume Assumption 1, i.e., there exists a one-to-one mapping between the parameter space and the joint distribution space. Parallel to the identification framework for the selection model, we must rule out values of θ that result in the identical distribution of observed data, which are characterized by

$$
\begin{aligned}
\mathrm{pr}(z; \theta_1) &= \mathrm{pr}(z; \theta_2) \\
\mathrm{pr}(y, r = 0 \mid z; \theta_1) &= \mathrm{pr}(y, r = 0 \mid z; \theta_2), \\
\mathrm{pr}(x, y, r = 1 \mid z; \theta_1) &= \mathrm{pr}(x, y, r = 1 \mid z; \theta_2).
\end{aligned}
$$

We have have the following condition for identification.

Condition A.1. *The parameter θ is identified, if for any two values θ_1 and θ_2 of θ such that* $\mathrm{pr}(z;\theta_1) = \mathrm{pr}(z;\theta_2)$ *and* $\mathrm{pr}(y, r = 0 \mid z;\theta_1) = \mathrm{pr}(y, r = 0 \mid z;\theta_2)$ *almost surely, the following inequality holds with a positive probability*

$$\frac{\mathrm{pr}(x, y \mid z, r = 0;\theta_1)}{\mathrm{pr}(x, y \mid z, r = 0;\theta_2)} \neq C \times \frac{\exp\{OR(x, y \mid z;\theta_1)\}}{\exp\{OR(x, y \mid z;\theta_2)\}}, \tag{1}$$

with

$$OR(x, y \mid z;\theta) = \log \frac{\mathrm{pr}(x, y \mid z, r = 0;\theta)\mathrm{pr}(x = 0, y \mid z, r = 1;\theta)}{\mathrm{pr}(x, y \mid z, r = 1;\theta)\mathrm{pr}(x = 0, y \mid z, r = 0;\theta)},$$

encoding the degree of departure between the two data patterns corresponding to $r = 0, 1$ respectively, and

$$C = \frac{E[\exp\{-OR(x, y \mid z;\theta_1)\} \mid r = 0]}{E[\exp\{-OR(x, y \mid z;\theta_2)\} \mid r = 0]}.$$

Condition A.1 is a sufficient condition for identification. One can verify that inequality (1) is in fact equivalent to $\mathrm{pr}(x, y, r = 1 \mid z;\theta_1) \neq \mathrm{pr}(x, y, r = 1 \mid z;\theta_2)$. However, (1) provides a useful access to check identification of the pattern-mixture parametrization where one specifies a parametric/semiparametric model for $\mathrm{pr}(x, y \mid z, r)$. In particular, when one has available a fully observed shadow variable z for the missing covariate x, i.e., $Z \perp\!\!\!\perp R \mid (X, Y)$, one can verify that

$$\mathrm{OR}(x, y \mid z;\theta) \quad = \quad \log \frac{\mathrm{pr}(r = 0 \mid x, y;\theta)\mathrm{pr}(r = 1 \mid x = 0, y;\theta)}{\mathrm{pr}(r = 1 \mid x, y;\theta)\mathrm{pr}(r = 0 \mid x = 0, y;\theta)},$$

which is a function only of (x, y). As a result, the right hand side of (1) does not vary with z. We have the following identification result for pattern-mixture model.

Proposition A.1. *Considering models* $\mathrm{pr}(y \mid x, z, r;\theta)$ *and* $\mathrm{pr}(x|z, r;\xi)$*, if for $\theta_1 \neq \theta_2$, the ratio* $\mathrm{pr}(x, y \mid z, r = 0;\theta_1, \xi_1)/\mathrm{pr}(x, y \mid z, r = 0;\theta_2, \xi_2)$ *varies with z for all ξ_1, ξ_2, then the parameter θ indexing the outcome model is identified.*

The proposition follows from the fact that under the shadow variable assumption, the right hand side of (1) is not a function of z, and thus (1) must hold if the ratio $\mathrm{pr}(x, y \mid z, r = 0; \theta_1, \xi_1)/\mathrm{pr}(x, y \mid z, r = 0; \theta_2, \xi_2)$ varies with z for distinct values θ_1 and θ_2. Assuming the generalized liner models (4)–(5) for $\mathrm{pr}(x \mid z, r = 0)$ and $\mathrm{pr}(y \mid x, z, r = 0)$ respectively, one can apply the results of Theorems 1–3 to check identification of pattern-mixture models.

B. Efficiency for (8)

We apply Newey and McFadden (1994, Theorem 5.3) to derive the optimal choice of G leading to the efficient estimator that solves (8). We let

$$U(G, \alpha) = \{r/\pi(x, y; \alpha) - 1\}\, G(z, y).$$

The IPW estimator $\widehat{\alpha}$ in this paper in fact solves $\widehat{E}\{U(G, \widehat{\alpha})\} = 0$. From Newey and McFadden (1994, Theorem 5.3), the optimal choice G_{opt} satisfies

$$E\left\{\partial U(G, \alpha_0)/\partial \alpha^T\right\} = E\{U(G, \alpha_0) U(G_{\mathrm{opt}}, \alpha_0)^T\}, \quad \text{for all } G(y, z),$$

with α_0 the true value of α. Thus,

$$E\left[G(y, z)\left\{(r/\pi(x, y; \alpha_0) - 1)^2 \times G_{\mathrm{opt}}^T + r/\pi^2(x, y; \alpha_0) \times \partial \pi(x, y; \alpha_0)/\partial \alpha^T\right\}\right] = 0,$$

for all $G(y, z)$. As a consequence, we have

$$E\left\{(r/\pi(x, y; \alpha_0) - 1)^2 \times G_{\mathrm{opt}} + r/\pi^2(x, y; \alpha_0) \times \partial \pi(x, y; \alpha_0)/\partial \alpha \mid y, z\right\} = 0,$$

and thus

$$G_{\mathrm{opt}}(y, z) = -1/E\{(r/\pi(x, y; \alpha_0) - 1)^2 \mid y, z\} \times E\left\{r/\pi^2(x, y; \alpha_0) \times \partial \pi(x, y; \alpha_0)/\partial \alpha \mid y, z\right\},$$

and the variance of the corresponding estimator is

$$
\begin{aligned}
V_{\text{opt}} &= [E\{U(G_{\text{opt}}, \alpha_0)U(G_{\text{opt}}, \alpha_0)^T\}]^{-1} \\
&= \left[E\left\{\{r/\pi(x, y, \alpha_0) - 1\}^2 \times G_{\text{opt}}(y, z)G_{\text{opt}}^T(y, z)\right\}\right]^{-1}.
\end{aligned}
$$

Under the shadow variable setting $Z \perp\!\!\!\perp R \mid (X, Y)$, we have

$$
\begin{aligned}
V_{\text{opt}} &= \left[E\left\{\{r/\pi(x, y, \alpha_0) - 1\}^2 \times G_{\text{opt}}(y, z)G_{\text{opt}}^T(y, z)\right\}\right]^{-1} \\
&= \left[E\left\{E\{(r/\pi(x, y, \alpha_0) - 1)^2 \mid x, y\} \times E\{G_{\text{opt}}(y, z)G_{\text{opt}}^T(y, z) \mid x, y\}\right\}\right]^{-1} \\
&= \left[E\left\{\{1/\pi(x, y, \alpha_0) - 1\} \times E\{G_{\text{opt}}(y, z)G_{\text{opt}}^T(y, z) \mid x, y\}\right\}\right]^{-1} \\
&= \left[E\left\{1/\pi(x, y, \alpha_0) - 1\right\} \times G_{\text{opt}}(y, z)G_{\text{opt}}^T(y, z)\}\right]^{-1}.
\end{aligned}
$$

The optimal choice G_{opt} and the variance V_{opt} depend the shadow variable Z. A choice of Z such that $E\left\{1/\pi(x, y, \alpha_0) - 1\right\} \times G_{\text{opt}}(y, z)G_{\text{opt}}^T(y, z)\}$ is large is desirable to maximize efficiency. Construction of G_{opt} depends on the unknown true data generating process and nuisance parameter $\text{pr}(x \mid y, z)$. A feasible approach is to plug-in consistent nuisance parameter estimates, but it is still difficult in particular for continuous y or z because $\text{pr}(x \mid y, z)$ may be very complicated.

C. Proofs of Theorems

We prove the identification results of Theorems 1–3 by verifying the condition of Proposition 1, i.e., the ratio $\text{pr}(y, x \mid z; \theta)/\text{pr}(y, x \mid z; \theta')$ varies with z for $\theta \neq \theta'$, which is determined by functions η_1, η_2, B_1, B_2 of models (4)–(5). We first describe four lemmas about these functions.

Lemma 1. *Suppose* $\text{pr}(x \mid z)$ *follows model* (4) *and* $(\gamma', \lambda') \neq (\gamma, \lambda)$*, then the ratio* $\text{pr}(x \mid z; \gamma', \lambda')/\text{pr}(x \mid z; \gamma, \lambda)$ *varies with* z*.*

Proof. The proof proceeds by contradiction. Suppose the ratio $\mathrm{pr}(x|z;\gamma',\lambda')/\mathrm{pr}(x|z;\gamma,\lambda)$ does not vary with z, and

$$\frac{\mathrm{pr}(x|z;\gamma',\lambda')}{\mathrm{pr}(x|z;\gamma,\lambda)} = h(x),$$

for some $h(x) \neq 1$, then we have

$$\int_x \mathrm{pr}(x \mid z;\gamma,\lambda)dx = \int_x \mathrm{pr}(x \mid z;\gamma',\lambda')dx = \int_x \mathrm{pr}(x \mid z;\gamma,\lambda)h(x)dx = 1,$$

for all z, and thus $\int_x \mathrm{pr}(x|z;\gamma,\lambda)\{h(x)-1\}dx = 0$ for all z, i.e.,

$$\int_x \exp\left\{\frac{x\cdot\eta_1(z;\gamma) - B_1(\eta_1(z;\gamma))}{\lambda} + A_1(x,\lambda)\right\}\{h(x)-1\}dx = 0, \qquad (2)$$

for all z. Under the full rank condition for the exponential family, X is complete for $\mathrm{pr}(x \mid z)$ (Shao, 2003, Proposition 2.1, page 110), i.e., $E\{f(X) \mid z\} = 0$ for all z implies $f(X) = 0$. Thus, from (2), we must have $h(x) = 1$, which contradicts $(\gamma',\lambda') \neq (\gamma,\lambda)$. As a result, $\mathrm{pr}(x \mid z;\gamma',\lambda')/\mathrm{pr}(x \mid z;\gamma,\lambda)$ must vary with z. □

Lemma 2. *Suppose the third order derivative function of B_2 denoted by $B_2^{(3)}$ is not a constant and let $g = B_2^{(3)}$. If $\beta^2 g(\alpha+\beta t) = \beta'^2 g(\alpha'+\beta' t)$ for all t, then we must have*

1. *$\beta = \beta'$; or*

2. *$\beta = -\beta' \neq 0$, and $g(\alpha+\beta t) = g(\alpha'-\beta t)$ for all t.*

Proof. If $\beta = 0$, $\beta'^2 g(\alpha'+\beta' t) = \beta^2 g(\alpha+\beta t) = 0$ for all t. Because g is a nonzero function, we must have $\beta' = 0$;

For $\beta \neq 0$, we must have $\beta' \neq 0$. For $|\beta'/\beta| < 1$, letting $s = \beta t$, because $\beta^2 g(\alpha + \beta t) = \beta'^2 g(\alpha'+\beta' t)$ for any t, we have

$$g(\alpha+s) = (\beta'/\beta)^2 \cdot g(\alpha' + \beta'/\beta \cdot s),$$

and thus

$$\begin{aligned} g(\alpha+s) &= (\beta'/\beta)^2 \cdot g(\alpha+(\alpha'-\alpha)+\beta'/\beta \cdot s) \\ &= (\beta'/\beta)^4 \cdot g(\alpha'+\beta'/\beta(\alpha'-\alpha)+\beta'^2/\beta^2 \cdot s). \end{aligned}$$

By iteration, we have $g(\alpha+s)=0$ for all s, which is impossible for a nonzero function g. So we have $|\beta'/\beta| \geq 1$, and similarly, $|\beta'/\beta| \leq 1$. As a result, we have $|\beta| = |\beta'| > 0$.

If $\beta = \beta' \neq 0$, we have $g(\alpha+\beta t) = g(\alpha'+\beta t)$ for all t. If $\beta = -\beta' \neq 0$, we have $g(\alpha+\beta t) = g(\alpha'-\beta t)$ for all t.

□

Lemma 3. *Suppose the first order derivative function of η_2 denoted by $\eta_2^{(1)}$ is not a constant and let $g=\eta_2^{(1)}$. For arbitrary $\phi, \phi' > 0$, if $\beta/\phi \cdot g(\alpha+\beta t) = \beta'/\phi' \cdot g(\alpha'+\beta' t)$ for all t, then we must have*

1. *$\beta = \beta'$; or*

2. *$\beta = -\beta' \neq 0$, $\phi = \phi'$, and $g(\alpha+\beta t) = -g(\alpha'-\beta t)$ for any t.*

Proof. We first prove that $|\beta'| \neq |\beta|$ is impossible by an argument of contradiction. Suppose $\beta \neq 0$. For $|\beta'/\beta| < 1$, because $\beta/\phi \cdot g(\alpha+\beta t) = \beta'/\phi' \cdot g(\alpha'+\beta' t)$ for any t, letting $s=\beta t$, we have $g(\alpha+s) = \beta'/\beta \cdot \phi/\phi' \cdot g(\alpha'+\beta'/\beta \cdot s)$. By iteration of the former formula, $\beta/\phi = \beta'/\phi'$ and $g(\alpha+s)$ must be a constant, which contradicts that $g=\eta_2^{(1)}$ is not a constant. Thus, $|\beta'/\beta| < 1$ is impossible and similarly $|\beta'/\beta| > 1$ is impossible. Thus, if $\beta \neq 0$, we must have $|\beta| = |\beta'|$. By switching (α, β, ϕ) and (α', β', ϕ') in the above argument, if $\beta' \neq 0$, we have $|\beta| = |\beta'|$. As a result, we have $|\beta'| = |\beta|$.

If further $\beta = -\beta' \neq 0$, we have $g(\alpha+\beta t) = -\phi'/\phi \cdot g(\alpha'-\beta t)$ for all t, and thus $g(\alpha'-\beta t) = -\phi'/\phi \cdot g(\alpha+\beta t)$ for all t. We let s_1 and s_2 denote two points such that

$g(s_1), g(s_2) \neq 0$, and let t_1, t_2 denote two values such that $\alpha' - \beta t_1 = \alpha + \beta t_2 = s_1$, and $\alpha' - \beta t_2 = \alpha + \beta t_1 = s_2$, then we have $g(s_1)/g(s_2) = g(\alpha' - \beta t_1)/g(\alpha + \beta t_1) = -\phi'/\phi$, and $g(s_1)/g(s_2) = g(\alpha + \beta t_2)/g(\alpha' - \beta t_2) = -\phi/\phi'$. As a result, we must have $\phi = \phi'$, and thus $g(\alpha + \beta t) = -g(\alpha' - \beta t)$ for all t.

$\square$

Lemma 4. *Suppose the third order derivative function of B_2 denoted by $B_2^{(3)}$ is not a constant and let $g = B_2^{(3)}$. If $g(\alpha + \beta t) = g(\alpha' + \beta' t)$ for all t, then we must have $|\beta| = |\beta'|$.*

Proof. We prove $|\beta| = |\beta'|$ by an argument of contradiction. Suppose $\beta \neq 0$, because $g(\alpha + \beta t) = g(\alpha' + \beta' t)$ for all t, by letting $s = \beta t$, we have

$$g(\alpha + s) = g(\alpha' + \beta'/\beta \cdot s), \quad \text{for all } s. \tag{3}$$

For $|\beta'/\beta| < 1$, by iteration of (3), we have $g(\alpha + s) = g\{\alpha + \sum_{k=0}^{+\infty} (\beta'/\beta)^k (\alpha' - \alpha)\}$. Thus, $g(\alpha + s)$ is a constant, which is a contradiction. Thus, $|\beta'/\beta| \leq 1$ is impossible, and similarly, $|\beta'/\beta| > 1$ is impossible. STherefore, if $\beta \neq 0$, we must have $|\beta| = |\beta'|$. By switching (α, β) and (α', β') in the above argument, if $\beta' \neq 0$, we have $|\beta| = |\beta'|$. In summary, we have $|\beta'| = |\beta|$.

$\square$

Proof of Theorem 1

According to Proposition 1, we prove the identification results of Theorem 1 by showing that the ratio $\mathrm{pr}(y, x|z; \theta)/\mathrm{pr}(y, x|z; \theta')$ varies with z when particular components of two different parameter sets θ and θ' are not equal. Letting $L(y, x, z) = \log\{\mathrm{pr}(y, x \mid$

$z;\theta)/\mathrm{pr}(y, x \mid z;\theta')\}$ and assuming models (4)–(5), we have

$$\begin{aligned} L(y,x,z) &= y \cdot \left(\frac{\eta_2}{\phi} - \frac{\eta_2'}{\phi'}\right) - \left\{\frac{B_2(\eta_2)}{\phi} - \frac{B_2(\eta_2')}{\phi'}\right\} + x \cdot \left\{\frac{\eta_1}{\lambda} - \frac{\eta_1'}{\lambda'}\right\} \\ &\quad - \left\{\frac{B_1(\eta_1)}{\lambda} - \frac{B_1(\eta_1')}{\lambda'}\right\} + \{A_2(y,\phi) - A_2(y,\phi')\} + \{A_1(x,\lambda) - A_1(x,\lambda')\}. \end{aligned}$$

(a) Letting

$$\frac{\partial^2 L}{\partial y \partial z} = \frac{\beta_1}{\phi}\eta_2^{(1)}(\beta_0 + \beta_1 z + \beta_2 x) - \frac{\beta_1'}{\phi'}\eta_2^{(1)}(\beta_0' + \beta_1' z + \beta_2' x),$$

if $\partial^2 L/(\partial y \partial z)$ is not equal to zero, then $L(y,x,z)$ varies with z. We prove identification of β_1/ϕ by showing that $\partial^2 L/(\partial y\partial z) \neq 0$ for $\beta_1/\phi \neq \beta_1'/\phi'$.

If η_2 is a linear function, i.e., $\eta_2^{(1)}$ is a nonzero constant, then $\partial^2 L/(\partial y \partial z)$ cannot equal zero for $\beta_1/\phi \neq \beta_1'/\phi'$. Thus, β_1/ϕ must be identified.

(b) We first prove identification under (i) $\beta_2 = \beta_2' = 0$. We then prove identification under (ii) $\beta_2 = \beta_2' = 0$ does not hold, by showing that $\partial^3 L/(\partial^2 x \partial z) \neq 0$ for $(\beta_1, \beta_2, \phi) \neq (\beta_1', \beta_2', \phi')$.

Under (i), we have $Y \perp\!\!\!\perp X \mid Z$, and thus $\mathrm{pr}(y \mid z, x) = \mathrm{pr}(y \mid z)$ can be identified from the observed data, thus, (β_1, β_2, ϕ) is identified.

Under (ii), we prove identification of (β_1, β_2, ϕ) by applying Lemmas 2 and 4 to show that $\partial^3 L/(\partial^2 x \partial z) \neq 0$ for $(\beta_1, \beta_2, \phi) \neq (\beta_1', \beta_2', \phi')$.

Because η_2 is a linear function, from (a) we have $\beta_1/\phi = \beta_1'/\phi'$ and

$$\frac{\partial^3 L}{\partial^2 x \partial z} = -\frac{\beta_1}{\phi}\{\beta_2^2 B_2^{(3)}(\beta_0 + \beta_1 z + \beta_2 x) - \beta_2'^2 B_2^{(3)}(\beta_0' + \beta_1' z + \beta_2' x)\}.$$

Because $B_2^{(2)}$ is a nonlinear function, $B_2^{(3)}$ is not a constant. We consider the following three cases for (ii).

(b1) If $|\beta_1| \neq |\beta_1'|$, from Lemma 2, $\partial^3 L/(\partial^2 x \partial z) \neq 0$.

(b2) If $\beta_2 = -\beta_2' \neq 0$, letting $z = -(\beta_0 + \beta_2 x)/\beta_1$, we have

$$\frac{\partial^3 L}{\partial^2 x \partial z} = -\frac{\beta_1}{\phi}\beta_2^2\{B_2^{(3)}(0) - B_2^{(3)}(\beta_0' - \beta_0 - 2\beta_2 x)\}.$$

Because $B_2^{(2)}$ is not a linear function, i.e., $B_2^{(3)}$ is not a constant, from Lemma 2, it is impossible that $\partial^3 L/(\partial^2 x \partial z) = 0$ for all x.

(b3) If $\beta_2 = \beta_2' \neq 0$ and $(\beta_1, \phi) \neq (\beta_1', \phi')$, we apply Lemma 4 to show $\partial^3 L/(\partial^2 x \partial z) \neq 0$. We have

$$\frac{\partial^3 L}{\partial^2 x \partial z} = -\frac{\beta_1 \beta_2^2}{\phi}\{B_2^{(3)}(\beta_0 + \beta_1 z + \beta_2 x) - B_2^{(3)}(\beta_0' + \beta_1' z + \beta_2 x)\}.$$

Because η_2 is a linear function, we have proved that $\beta_1/\phi = \beta_1'/\phi'$ in (a). Because $\phi, \phi' > 0$, β_1 and β_1' must have the same sign. For fixed x, from Lemma 4, $\partial^3 L/(\partial^2 x \partial z) \neq 0$ for $\beta_1 \neq \beta_1'$ or $\phi \neq \phi'$.

From (b1)–(b3), we have shown that under (ii), $\partial^3 L/(\partial^2 x \partial z) \neq 0$ for $(\beta_1, \beta_2, \phi) \neq (\beta_1', \beta_2', \phi')$. Thus, applying Proposition 1, (β_1, β_2, ϕ) must be identified under (ii).

Therefore, we have proved that when η_2 is a linear function and $B_2^{(2)}$ is a nonlinear function, (β_1, β_2, ϕ) are identified.

(c) We first prove identification under (i) $\beta_1 = \beta_1' = 0$. We then prove identification when (ii) $\beta_1 = \beta_1' = 0$ does not hold, by showing that $\partial^2 L/(\partial y \partial z) \neq 0$ for $(\beta_1, \beta_2, \phi) \neq (\beta_1', \beta_2', \phi')$.

Under (i) $\beta_1 = \beta_1' = 0$, we have $Y \perp\!\!\!\perp Z \mid X$. Noting the shadow variable assumption $Z \perp\!\!\!\perp R \mid (Y, X)$, we have $Z \perp\!\!\!\perp R \mid X$, and thus

$$L(y, x, z) = \log \frac{\mathrm{pr}(x \mid z; \gamma, \lambda)}{\mathrm{pr}(x \mid z; \gamma', \lambda')} + \log \frac{\mathrm{pr}(y \mid x; \beta_2, \phi)}{\mathrm{pr}(y \mid x; \beta_2', \phi')}.$$

If $(\gamma, \lambda) \neq (\gamma', \lambda')$, from Lemma 1, $\mathrm{pr}(x \mid z; \gamma, \lambda)/\mathrm{pr}(x \mid z; \gamma', \lambda')$ varies with z, and so does $L(y, x, z)$.

If $(\gamma, \lambda) = (\gamma', \lambda')$, we note that $\mathrm{pr}(y \mid z)$ is identified and

$$\mathrm{pr}(y \mid z) = \int_x \mathrm{pr}(x|z; \gamma, \lambda)\mathrm{pr}(y \mid x; \beta_2, \phi)dx = \int_x \mathrm{pr}(x \mid z)\mathrm{pr}(y \mid x; \beta_2', \phi')dx,$$

i.e., for all z, we have the following integral equation

$$\int_x \exp\left\{x \cdot \frac{\eta_1(z;\gamma)}{\lambda} - B_1(\eta_1(z;\gamma)) + A_1(x;\lambda)\right\}\{\mathrm{pr}(y \mid x; \beta_2, \phi) - \mathrm{pr}(y \mid x; \beta_2', \phi')\}dx = 0,$$

thus, by completeness of the exponential families under the full rank condition (Shao, 2003, Proposition 2.1, page 110), we have $\mathrm{pr}(y \mid x; \beta_2, \phi) = \mathrm{pr}(y \mid x; \beta_2', \phi')$.

As a result, we have shown identification of (β_1, β_2, ϕ) under (i).

Under (ii), we apply Lemma 3 to prove identification of (β_1, β_2) by showing that $\partial^2 L/(\partial y \partial z) \neq 0$ for $(\beta_1, \beta_2) \neq (\beta_1', \beta_2')$. We consider the following three cases.

(c1) Because η_2 is a nonlinear function, $\eta_2^{(1)}$ is not a constant. If $|\beta_1| \neq |\beta_1'|$, then from Lemma 3, $\partial^2 L/(\partial y \partial z) \neq 0$.

(c2) If $\beta_1 = -\beta_1' \neq 0$, we show that $\partial^2 L/(\partial y \partial z)$ cannot equal zero for all x.

If $\beta_1 = -\beta_1' \neq 0$ and $\phi \neq \phi'$, from Lemma 3, $\partial^2 L/(\partial y \partial z)$ cannot equal zero for all x.

If $\beta_1 = -\beta_1' \neq 0$, $\phi = \phi'$, and $\beta_2 \neq -\beta_2'$, letting $z = -(\beta_0 + \beta_2 x)/\beta_1$, we have

$$\frac{\partial^2 L}{\partial y \partial z} = \frac{\beta_1}{\phi}[\eta_2^{(1)}(0) + \eta_2^{(1)}\{\beta_0 + \beta_0' + (\beta_2 + \beta_2')x\}],$$

which cannot equal 0 for all x because $\eta_2^{(1)}$ is not a constant.

If $\beta_1 = -\beta_1' \neq 0$ and $(\phi, \beta_2) = (\phi', -\beta_2')$, we let $g(x, z) = \eta_2(\beta_0 + \beta_1 z + \beta_2 x) - \eta_2(\beta_0' - \beta_1 z - \beta_2 x)$. If $\partial g(x, z)/\partial z \neq 0$, we have $\partial^2 L/(\partial y \partial z) = \beta_1/\phi \times$

$\partial g(x,z)/\partial z \neq 0$; otherwise if $\partial g(x,z)/\partial z = 0$, i.e., $g(z,x) = g(x)$ is a function only of x, we let $z = (\beta_0' - \beta_0 - 2\beta_2 x)/(2\beta_1)$ and then we must have $g(x) = \eta_2\{(\beta_0+\beta_0')/2\} - \eta_2\{(\beta_0+\beta_0')/2\} = 0$ for all x. Therefore, $\eta_2(\beta_0+\beta_1 z+\beta_2 x) = \eta_2(\beta_0' - \beta_1 z - \beta_2 x)$ for all z and for all x. Note that $\phi = \phi'$, then the two different sets $(\beta_0, \beta_1, \beta_2, \phi)$ and $(\beta_0', \beta_1', \beta_2', \phi')$ must index the identical distribution $\mathrm{pr}(y \mid z, x)$, which contradicts Assumption 1 that we assume a one-to-one mapping between parameters and the joint distribution.

As a result, if $\beta_1 = -\beta_1' \neq 0$, $\partial^2 L/(\partial y \partial z)$ cannot equal zero for all x.

(c3) If $\beta_1 = \beta_1' \neq 0$ and $\beta_2 \neq \beta_2'$, we show $\partial^2 L/(\partial y \partial z) \neq 0$. For $\beta_1 = \beta_1' \neq 0$, we have

$$\frac{\partial^2 L}{\partial y \partial z} = \beta_1 \left\{ \frac{1}{\phi}\eta_2^{(1)}(\beta_0 + \beta_1 z + \beta_2 x) - \frac{1}{\phi'}\eta_2^{(1)}(\beta_0' + \beta_1 z + \beta_2' x) \right\}.$$

Letting $z = -(\beta_0 + \beta_2 x)/\beta_1$, we have

$$\frac{\partial^2 L}{\partial y \partial z} = \beta_1 \left[\frac{1}{\phi}\eta_2^{(1)}(0) - \frac{1}{\phi'}\eta_2^{(1)}\{\beta_0' - \beta_0 + (\beta_2' - \beta_2)x\} \right],$$

which cannot equal 0 for all x because $\eta_2^{(1)}$ is not a constant. Thus, $\partial^2 L/(\partial y \partial z) \neq 0$.

From (c1)–(c3), we have shown that under (ii), $\partial L/(\partial y \partial z) \neq 0$ for $(\beta_1, \beta_2) \neq (\beta_1', \beta_2')$. Thus, applying Proposition 1, (β_1, β_2) must be identified under (ii). Therefore, we have proved that when η_2 is a non-linear function, (β_1, β_2) are identified.

Proof of Theorem 2

Assume the normal models: $Y \mid X, Z \sim N(\beta_0 + \beta_1 z + \beta_2 x, \phi)$ and $X \mid Z \sim$

$N(\gamma_0 + \gamma_1 z, \lambda)$, then we have the following conditional distribution

$$X \mid Y, Z \sim N(\gamma_0' + \gamma_1' z + \gamma_2' y, \lambda'),$$

with

$$\gamma_0' = \gamma_0 - \frac{\beta_2\lambda(\beta_0 + \beta_2\gamma_0)}{\phi + \beta_2^2\lambda}, \quad \gamma_1' = \gamma_1 - \frac{\beta_2\lambda(\beta_1 + \beta_2\gamma_1)}{\phi + \beta_2^2\lambda},$$
$$\gamma_2' = \frac{\beta_2\lambda}{\phi + \beta_2^2\lambda}, \quad \lambda' = \frac{\phi\lambda}{\phi + \beta_2^2\lambda}.$$

Because $X \perp\!\!\!\perp Z \mid Y$ if and only if $\gamma_1' = 0$, the shadow variable assumption is satisfied when $\gamma_1' \neq 0$, i.e., $\beta_1\beta_2/\phi \neq \gamma_1/\lambda$. Under such a condition, because $\mathrm{pr}(x \mid y, z)$ follows a normal model, Miao et al. (2015) proved that for any two candidate models $\mathrm{pr}(x \mid y, z)$ and $\mathrm{pr}'(x \mid y, z)$, the ratio $\mathrm{pr}(x \mid y, z)/\mathrm{pr}'(x \mid y, z)$ must vary with z. Thus, $\mathrm{pr}(x, y, z)/\mathrm{pr}'(x, y, z)$ must vary with z, and therefore, all parameters $(\beta_0, \beta_1, \beta_2, \phi, \lambda, \alpha_0, \alpha_1, \alpha_2)$ are identified.

Proof of Theorem 3

(a) If $\beta_1 = \beta_1' = 0$, i.e., $Y \perp\!\!\!\perp Z \mid X$, then from the shadow variable assumption $Z \perp\!\!\!\perp R \mid (Y, X)$, we have $Z \perp\!\!\!\perp R \mid X$, and thus

$$L(y, x, z) = \log \frac{\mathrm{pr}(x \mid z; \gamma, \lambda)}{\mathrm{pr}(x \mid z; \gamma', \lambda')} + \log \frac{\mathrm{pr}(y \mid x; \beta_2, \phi)}{\mathrm{pr}(y \mid x; \beta_2', \phi')}.$$

If $(\gamma, \lambda) \neq (\gamma', \lambda')$, from Lemma 1, $\mathrm{pr}(x \mid z; \gamma, \lambda)/\mathrm{pr}(x \mid z; \gamma', \lambda')$ varies with z, and so does $L(y, x, z)$.

If $(\gamma, \lambda) = (\gamma', \lambda')$, we note that $\mathrm{pr}(y \mid z)$ is identified and

$$\mathrm{pr}(y \mid z) = \int_x \mathrm{pr}(x|z; \gamma, \lambda)\mathrm{pr}(y \mid x; \beta_2, \phi)dx = \int_x \mathrm{pr}(x \mid z)\mathrm{pr}(y \mid x; \beta_2', \phi')dx,$$

i.e., for all z, we have the following integral equation

$$\int_x \exp\left\{x \cdot \frac{\eta_1(z; \gamma)}{\lambda} - B_1(\eta_1(z; \gamma)) + A_1(x; \lambda)\right\}\{\mathrm{pr}(y \mid x; \beta_2, \phi) - \mathrm{pr}(y \mid x; \beta_2', \phi')\}dx = 0,$$

thus, by completeness of the exponential families under the full rank condition (Shao, 2003, Proposition 2.1, page 110), we have $\mathrm{pr}(y \mid x; \beta_2, \phi) = \mathrm{pr}(y \mid x; \beta_2', \phi')$. As a result, we have shown identification of (β_2, ϕ) for model 6.

(b) If $\beta_2 = \beta_2' = 0$, we have $Y \perp\!\!\!\perp X | Z$, and thus $\mathrm{pr}(y \mid z, x) = \mathrm{pr}(y \mid z)$, which can be identified from the observed data. As a result, (β_0, β_1, ϕ) are identified under model 7.

Proof of Theorem 4

We prove that (8) and (9) are unbiased estimating equations, when both $\mathrm{pr}(r = 1 \mid x, y; \alpha)$ and $\mathrm{pr}(y \mid x, z; \beta)$ are correctly specified. Under the shadow variable assumption $Z \perp\!\!\!\perp R \mid (X, Y)$, at the true value α^0 of α, we have

$$\begin{aligned} E\left[\left\{\frac{r}{\pi(x, y; \alpha^0)} - 1\right\} G(z, y)\right] &= E\left[E\left\{\left(\frac{r}{\pi(x, y; \alpha^0)} - 1\right) G(z, y) \mid x, y\right\}\right] \\ &= E\left[E\left\{\frac{r}{\pi(x, y; \alpha^0)} - 1 \mid x, y\right\} \times E\left\{G(z, y) \mid x, y\right\}\right]. \end{aligned}$$

When $\mathrm{pr}(r = 1 \mid x, y; \alpha)$ is correctly specified, $E\{r/\pi(x, y; \alpha^0) - 1 \mid x, y\} = 0$, and thus $E\left[\left\{r/\pi(x, y; \alpha^0) - 1\right\} G(z, y)\right] = 0$, i.e., (8) is an unbiased estimating equation for α. Furthermore, under true values $(\alpha^0, \beta^0, \phi^0)$, we have

$$\begin{aligned} E\left\{\frac{r}{\pi(x, y; \alpha)} S(x, y; \beta, \phi)\right\} &= E\left\{E\left(\frac{r}{\pi(x, y; \alpha)} \mid x, y\right) \times S(x, y; \beta, \phi)\right\} \\ &= E\left\{S(x, y; \beta, \phi)\right\}, \end{aligned}$$

which equals zero under correct specification of both $\mathrm{pr}(r = 1 \mid x, y; \alpha)$ and $\mathrm{pr}(y \mid x, z; \beta)$. Thus, (9) is an unbiased estimating equation for (β, ϕ).

References

Miao, W., E. Tchetgen Tchetgen, and Z. Geng (2015). Identification and doubly robust estimation of data missing not at random with a shadow variable. Technical report.

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

Shao, J. (2003). *Mathematical Statistics* (2nd ed.). New York: Springer.

Wang Miao

Department of Business Statistics and Econometrics

Peking University

Haidian District, Beijing 100871

E-mail: mwfy@pku.edu.cn

Eric Tchetgen Tchetgen

Department of Biostatistics

Harvard University

Boston, Massachusetts 02115

E-mail: etchetge@hsph.harvard.edu