Semiparametric Estimation with Data Missing Not at Random Using an Instrumental Variable

BaoLuo Sun[1], Lan Liu[2], Wang Miao[3], Kathleen Wirth[4,5],

James Robins[1,4] and Eric J. Tchetgen Tchetgen[1,4]

Departments of Biostatistics[1], Epidemiology[4] and Immunology and Infectious Diseases[5], Harvard T.H. Chan School of Public Health.

School of Statistics[2], University of Minnesota at Twin Cities.

Department of Business Statistics & Econometrics[3], Guanghua School of Management, Peking University.

Supplementary Material

This supplement contains proofs for the Theorems, Propositions and Examples in the main manuscript.

S1 Proof of Result 1

The proof is based on contradiction. By the exclusion restriction assumption **(IV.1)** the decomposition of the joint distribution for (Z, Y, R) is

$$P_{\theta_i,\eta_i,\xi_i}(z, y, r) = P_{\theta_i}(r|z, y)P_{\eta_i}(z)P_{\xi_i}(y), \quad i = 1, 2, ..., n$$

Suppose we have two sets of candidates satisfying the same observed quantities:

$$P_{\theta_1}(z, y, R = 1) = P_{\theta_2}(z, y, R = 1)$$

$$P_{\eta_1}(z) = P_{\eta_2}(z)$$

Substituting the above observed quantities into the joint distribution gives

$$\frac{P_{\theta_1}(R = 1|z, y)}{P_{\theta_2}(R = 1|z, y)} = \frac{P_{\xi_2}(y)}{P_{\xi_1}(y)}$$

This contradicts with the requirement that the ratios are unequal.

S2 Proofs of Examples 1 and 2

Proof of Example 1

For binary outcome Y and binary instrument Z, let $P(R = 1|Z, Y; \theta) = \text{expit}\,[\theta_0 + \theta_1 Z + \theta_2 Y + \theta_3 ZY]$ and $P(Y = 1; \xi) = \exp(\xi)$. We show that for every (θ, ξ), there exists $(\tilde{\theta}, \tilde{\xi}) \neq (\theta, \xi)$ such that

$$\frac{P(R = 1|Z, Y; \theta)}{P(R = 1|Z, Y; \tilde{\theta})} = \frac{P(Y; \tilde{\xi})}{P(Y; \xi)} \tag{A}$$

Let $\frac{P(Y=0;\tilde{\xi})}{P(Y=0;\xi)} = \exp(\rho_0)$ for some $\rho_0 \neq 0$, then $\frac{P(Y;\tilde{\xi})}{P(Y;\xi)} = \exp(\rho_0 + \rho_1 Y)$ where

$$\rho_1 = \log\left\{\exp(-\rho_0 - \xi) + [\exp(\xi) - 1]/\exp(\xi)\right\}.$$

Equality (A) then holds by choosing $(\tilde{\theta}, \tilde{\xi})$ such that

$$\tilde{\theta}_0 = \theta_0 - \rho_0 - \log(\alpha_0)$$

$$\tilde{\theta}_1 = \theta_1 + \log(\alpha_0) - \log(\alpha_1)$$

$$\tilde{\theta}_2 = \theta_2 - \rho_1 + \log(\alpha_0) - \log(\alpha_2)$$

$$\tilde{\theta}_3 = \theta_3 + \log(\alpha_1) + \log(\alpha_2) - \log(\alpha_0) - \log(\alpha_3)$$

$$\tilde{\xi} = \xi + \rho_0 + \rho_1,$$

where $\alpha_0 = 1+\exp(\theta_0)-\exp(\theta_0-\rho_0)$, $\alpha_1 = 1+\exp(\theta_0+\theta_1)-\exp(\theta_0+\theta_1-\rho_0)$, $\alpha_2 = 1+\exp(\theta_0+\theta_2)-\exp(\theta_0+\theta_2-\rho_0-\rho_1)$ and $\alpha_3 = 1+\exp(\theta_0+\theta_1+\theta_2+\theta_3)-\exp(\theta_0+\theta_1+\theta_2+\theta_3-\rho_0-\rho_1)$. For example, choose $(\rho_0, \rho_1) = (0.3, -0.38)$ and equality (A) holds for $(\theta_0, \theta_1, \theta_2, \theta_3, \xi) = (0.3, 0.6, 0.1, 0.7, -0.2)$ and $(\tilde{\theta}_0, \tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3, \tilde{\xi}) = (-0.3, 0.41, 0.91, 1.37, -0.28)$.

Next, we consider the missingness mechanism $P(R = 1|Z, Y; \theta) = \text{expit}\,[\theta_0 + \theta_1 Z + \theta_2 Y]$, where the interaction effect between (Z, Y) is absent. Under this mechanism, we have $\theta_3 = \tilde{\theta}_3 = 0$ and therefore $\alpha_1\alpha_2 = \alpha_0\alpha_3$ which implies the equality

$$\exp(\rho_0 + \rho_1) = \frac{\exp(\theta_2 + \rho_0)}{\exp(\theta_2 + \rho_0) + [1 - \exp(\rho_0)]}. \tag{B}$$

Since $\exp(\rho_0+\rho_1 Y)$ is the ratio of the two probability mass distributions for Y, ρ_0 and $\rho_0 + \rho_1$ should be of opposite signs. Based on (B), if $\exp(\rho_0) > 1$ then $\exp(\rho_0 + \rho_1) > 1$ and similarly if $\exp(\rho_0) < 1$ then $\exp(\rho_0 + \rho_1) < 1$, which implies that the only possibility is $\rho_0 = \rho_1 = 0$ and hence $(\tilde{\theta}, \tilde{\xi}) = (\theta, \xi)$.

Proof of Example 2

Consider the case where Z and Y are both continuous random variables. Suppose two sets of candidates in the separable logistic missing data mechanism has the following relationship

$$\frac{\text{expit}(q_1(z) + h_1(y))}{\text{expit}(q_2(z) + h_2(y))} = g(y)$$

for some function $g(\cdot)$, i.e. the ratio is a function of y only. Taking derivative with respect to Z on both sides (assuming IV relevance **(IV.2)** holds) gives

$$\frac{\frac{\partial}{\partial z}\text{expit}(q_1(z) + h_1(y))}{\text{expit}(q_1(z) + h_1(y))} = \frac{\frac{\partial}{\partial z}\text{expit}(q_2(z) + h_2(y))}{\text{expit}(q_2(z) + h_2(y))}$$

or equivalently

$$\frac{\partial q_1(z)/\partial z}{\partial q_2(z)/\partial z} = \frac{1 + \exp(q_1(z) + h_1(y))}{1 + \exp(q_2(z) + h_2(y))} \tag{A}$$

Taking derivatives with respect to Y on both sides leads to

$$\frac{\partial q_1(z)/\partial z}{\partial q_2(z)/\partial z}\exp(q_2(z) - q_1(z)) = \frac{\partial h_1(y)/\partial y}{\partial h_2(y)/\partial y}\exp(h_1(y) - h_2(y))$$

The left hand side of the above equation depends only on Z but the right hand side depends only on Y, so it must be that

$$\frac{\partial q_1(z)/\partial z}{\partial q_2(z)/\partial z}\exp(q_2(z) - q_1(z)) = c_1$$

for some constant c_1. Substituting the above expression into equality (A) leads to

$$c_1 \left\{\exp(-q_2(z)) + \exp(h_2(y))\right\} = \exp(-q_1(z)) + \exp(h_1(y))$$

and therefore

$$c_1 \exp(-q_2(z)) + c_2 = \exp((-q_1(z)), \quad c_1 \exp(h_2(y)) - c_2 = \exp((h_1(y))$$

for some constant c_2. Substituting the above equalities into the ratio of propensity scores

$$\frac{\text{expit}(q_1(z) + h_1(y))}{\text{expit}(q_2(z) + h_2(y))} = 1 + c_2 \exp(-h_1(y)) = g(y)$$

Note that $g(y)$ is the ratio of two candidate densities of Y, and so it must be that $c_2 = 0$ and the two sets of candidates are equivalent, leading to a contradiction. Therefore the ratio

$$\frac{\text{expit}(q_1(z) + h_1(y))}{\text{expit}(q_2(z) + h_2(y))}$$

is either a constant or depends on z, which by Corollary 1 leads to identifiability of this class of missing data models.

Consider the case where Z is a binary random variable, and assume two sets of candidates in the separable logistic missing data mechanism has the following relationship

$$\frac{\text{expit}(\eta_1 z + h_1(y))}{\text{expit}(\eta_2 z + h_2(y))} = g(y).$$

The above relationship holds for $z = 0, 1$, therefore

$$\frac{\text{expit}(h_1(y))}{\text{expit}(h_2(y))} = \frac{\text{expit}(\eta_1 + h_1(y))}{\text{expit}(\eta_2 + h_2(y))}$$

and

$$g(y) = 1 + \frac{\exp(\eta_2) - \exp(\eta_1)}{\exp(\eta_2) - \exp(\eta_1 + \eta_2)} \exp[-h_2(y)].$$

Since $g(y)$ is the ratio of two densities, we must have $\eta_1 = \eta_2$ and $g(y) = 1$, leading to a contradiction. The proof for Y or Z as discrete variables is similar to the above proof for binary Z.

S3 Proofs of Propositions

Proof of Proposition 1

Let $(\eta_0, \omega_0, \xi_0)$ denote the true values of the parameters for parametric models $\eta(x, y, z; \zeta), P(r|Y = 0, x, z; \omega)$ and $q(z|x; \xi)$ which are assumed to be correctly specified. Assume the model $q(z|x; \xi)$ is identifiable, its parameter space Ξ is compact and the remaining conditions in Theorem 2.5 of Newey and McFadden (1994) hold, which are sufficient to establish consistency of maximum likelihood estimators. Then $\hat{\xi}_{\text{MLE}}$ has a probability limit equal to ξ_0. Consider estimating function for (4.6) which under the law of iterated expectations equals to

$$
\begin{aligned}
&E\left\{E\left\{\left[\frac{R}{\pi(\zeta_0,\omega_0)}-1\right]\boldsymbol{h}_1(X,Z)\right\}\middle|X,Y,Z\right\}\\
=&E\left\{E\left\{\left[\frac{\pi(\zeta_0,\omega_0)}{\pi(\zeta_0,\omega_0)}-1\right]\boldsymbol{h}_1(X,Z)\right\}\right\}=0.
\end{aligned}
$$

Under the law of iterated expectations, the estimating function for (4.7) equals

$$
\begin{aligned}
&E\left\{\frac{R}{\pi(\zeta_0,\omega_0)}g(Y,X)\{h_2(Z,X)-E[h_2(Z,X)|X;\xi_0]\}\right\}\\
=&E\left\{g(Y,X)\{h_2(Z,X)-E[h_2(Z,X)|X;\xi_0]\}\right\}\\
=&E\left\{E[g(Y,X)|X]\{h_2(Z,X)-E[h_2(Z,X)|X;\xi_0]\}\right\} \quad \text{by } \textbf{(IV.1)}\\
=&E\left\{E[g(Y,X)|X]\{E[h_2(Z,X)|X;\xi_0]-E[h_2(Z,X)|X;\xi_0]\}\right\}. = \quad 0.
\end{aligned}
$$

Therefore (η_0,ω_0) are the probability limits of the solutions to estimating equations (4.6) and (4.7). The IPW estimator is also unbiased,

$$
E\left\{\frac{RY}{\pi(\zeta_0,\omega_0)}\right\}=E\{Y\}=\phi_0,
$$

by taking iterated expectations with respect to (X,Y,Z). The consistency and asymptotic normality of $\hat{\phi}^{\text{IPW}}$ can be established under standard regularity conditions for GMM estimators (Newey and McFadden (1994)) , typically by placing moment restrictions on the vector of estimating functions. In particular, we require that the probability of observing the outcome is bounded away from zero, a necessary assumption for identification of a full data functional (Robins et al. (1994)).

$$\pi(x,y,z) > \sigma > 0 \quad \text{with probability } 1 \tag{S1}$$

for a non-zero positive constant $\sigma > 0$.

Let $\boldsymbol{M}(\delta)$ represent the stacked vector of the following estimating functions: score functions for estimating ξ, $\boldsymbol{U}^{\text{IPW}}(\xi,\zeta,\omega)$ and $G(\phi,\zeta,\omega)$ $= \left\{\frac{RY}{\pi(\zeta,\omega)} - \phi\right\}$, where $\delta = (\zeta,\omega,\xi,\phi)$. Then under standard regularity conditions for M-estimation (Newey and McFadden (1994)), the asymptotic variance V_{IPW} is given by the diagonal entry corresponding to ϕ of the following variance-covariance matrix

$$\left[E\left\{\frac{\partial \boldsymbol{M}(\delta)}{\partial \delta^T}\bigg|_{\delta_0}\right\}\right]^{-1} E\left\{\boldsymbol{M}(\delta_0)\boldsymbol{M}(\delta_0)^T\right\}\left[E\left\{\frac{\partial \boldsymbol{M}(\delta)}{\partial \delta^T}\right\}\bigg|_{\delta_0}\right]^{-1^T}, \tag{S2}$$

where $\delta_0 = (\zeta_0,\omega_0,\xi_0,\phi_0)$ is the probability limit of $\hat{\delta} = (\hat{\zeta},\hat{\omega},\hat{\xi},\hat{\phi})$. A consistent sandwich estimator for the above asymptotic variance can be constructed by evaluating unknown expectations as sample means at the estimated parameter value $\hat{\delta}$.

Proof of Proposition 2

Let (η_0,θ_0,ξ_0) denote the true values of the parameters for parametric models $\eta(x,y,z;\zeta), f(y|R=1,x,z;\theta)$ and $q(z|x;\xi)$ which are assumed to be correctly specified. Assume the conditions in Theorem 2.5 of Newey and McFadden (1994) hold for models $f(y|R=1,x,z;\theta)$ and $q(z|x;\xi)$.

Then the probability limits of the MLEs $(\hat{\theta}_{\text{MLE}}, \hat{\xi}_{\text{MLE}})$ are (θ_0, ξ_0). Under true parameter values, the expectation of the estimating function for (4.9) is

$$
\begin{aligned}
& E\Big\{\{q_1(X,Z) - E\left[q_1(X,Z)|X;\xi_0\right]\} \times \\
& \{(1-R)E\left(q_2(X,Y)|R=0,X,Z;\zeta_0,\theta_0\right) + Rq_2(X,Y)\}\Big\} \\
=& E\{E(\cdot|R=0,X,Z) \times \Pr(R=0|X,Z)\} \\
& + E\{E(\cdot|R=1,X,Z) \times \Pr(R=1|X,Z)\} \\
=& E\left(\{q_1(X,Z) - E[q_1(X,Z)|X;\xi_0]\}\, E[q_2(X,Y)|X,Z]\right) \\
=& E\left(\{q_1(X,Z) - E[q_1(X,Z)|X;\xi_0]\}\, E[q_2(X,Y)|X]\right) \qquad \text{by } \textbf{(IV.1)} \\
=& E\left(\{E[q_1(X,Z)|X;\xi_0] - E[[q_1(X,Z)|X;\xi_0]\}\, E[q_2(X,Y)|X]\right) \\
=& 0,
\end{aligned}
$$

so that ζ_0 is the probability limit of the solution $\hat{\zeta}$ of (4.9). The OR estimator is unbiased since

$$
\begin{aligned}
& E\left\{RY+(1-R)E(Y|R=0,X,Z;\zeta_0,\theta_0)\right\} \\
=& E\left\{E\{RY+(1-R)E(Y|R=0,X,Z)|R=0,X,Z\}\times \Pr(R=0|X,Z)\right\} \\
+& E\left\{E\{RY+(1-R)E(Y|R=0,X,Z)|R=1,X,Z\}\times \Pr(R=1|X,Z)\right\} \\
=& E\left\{E\{Y|R=0,X,Z\}\times \Pr(R=0|X,Z)\right\} \\
& + E\left\{E\{Y|R=1,X,Z\}\times \Pr(R=1|X,Z)\right\} \\
=& E\left\{E\{Y|X,Z\}\right\} \\
=& E\{Y\}=\phi_0.
\end{aligned}
$$

The consistency and asymptotic normality of $\hat{\phi}^{\mathrm{OR}}$ can be established under standard regularity conditions for GMM estimators (Newey and McFadden (1994)). A necessary condition is that the probability of observing the outcome is bounded away from zero (S1). Let $\boldsymbol{M}(\delta)$ represent the stacked vector of the following estimating functions: score functions for estimating ξ and θ, $\boldsymbol{U}^{\mathrm{OR}}(\xi,\zeta,\theta)$ and

$$G(\phi,\zeta,\theta)=\{RY+(1-R)E(Y|R=0,X,Z;\zeta,\theta)-\phi\},$$

where $\delta=(\zeta,\theta,\xi,\phi)$. Then under standard regularity conditions for M-estimation (Newey and McFadden (1994)), the asymptotic variance V_{OR} is given by the diagonal entry corresponding to ϕ of the following variance-covariance matrix

$$\left[E\left\{\frac{\partial \boldsymbol{M}(\delta)}{\partial \delta^T}\bigg|_{\delta_0}\right\}\right]^{-1} E\left\{\boldsymbol{M}(\delta_0)\boldsymbol{M}(\delta_0)^T\right\}\left[E\left\{\frac{\partial \boldsymbol{M}(\delta)}{\partial \delta^T}\right\}\bigg|_{\delta_0}\right]^{-1^T}, \quad \text{(S3)}$$

where $\delta_0 = (\zeta_0, \theta_0, \xi_0, \phi_0)$ is the probability limit of $\hat{\delta} = (\hat{\zeta}, \hat{\theta}, \hat{\xi}, \hat{\phi})$. A consistent sandwich estimator for the above asymptotic variance can be constructed by evaluating unknown expectations as sample means at the estimated parameter value $\hat{\delta}$.

Proof of Proposition 3

Under model $\mathcal{M}_{\text{IPW}}$, let ξ_0 denote the true value for parametric model $q(z|x;\xi)$ and it is clear that $\hat{\xi}_{\text{MLE}}$ has a probability limit equal to ξ_0. Let superscript asterisks denote possibly misspecified models. Let θ^* denote the probability limit of estimation under model $f^*(y|R=1,x,z;\theta)$ and let $\rho(X,Z) = \int \boldsymbol{u}(x,y)\frac{\exp[-\eta(x,y,z;\zeta)]f(y|R=1,x,z;\theta)}{\int \exp[-\eta(x,y,z)]f(y|R=1,x,z;\theta)\mathrm{d}\mu(y)}\mathrm{d}\mu(y)$. Then at true parameter values (ζ_0, ω_0),

$$\begin{aligned}&E\left\{\boldsymbol{G}^{\text{DR}}\left(R,X,Y,Z;\zeta_0,\omega_0,\theta^*,\boldsymbol{u}\right)|X,Y,Z\right\}\\=&\boldsymbol{u}(X,Y)-\rho^*(X,Z;\zeta_0,\theta^*)+\rho^*(X,Z;\zeta_0,\theta^*)=\boldsymbol{u}(X,Y),\end{aligned}$$

and therefore the estimating function for (4.12), under iterated expectations with respect to (X, Y, Z) at $(\xi_0, \zeta_0, \omega_0)$, is

$$
\begin{aligned}
&E\Bigg\{ [\boldsymbol{v}(X,Z) - E(\boldsymbol{v}(X,Z)|X)] \{\boldsymbol{u}(X,Y)\} \Bigg\} \\
=&E\Bigg\{ [\boldsymbol{v}(X,Z) - E(\boldsymbol{v}(X,Z)|X)] \{E(\boldsymbol{u}(X,Y)|X,Z)\} \Bigg\} \\
=&E\Bigg\{ [\boldsymbol{v}(X,Z) - E(\boldsymbol{v}(X,Z)|X)] \{E(\boldsymbol{u}(X,Y)|X)\} \Bigg\} \qquad \text{by } \textbf{(IV.1)} \\
=&E\Bigg\{ [E(\boldsymbol{v}(X,Z)|X) - E(\boldsymbol{v}(X,Z)|X)] \{E(\boldsymbol{u}(X,Y)|X)\} \Bigg\} \\
=&\mathbf{0}.
\end{aligned}
$$

In addtion, under iterated expectations with respect to (X, Y, Z),

$$E\{\boldsymbol{G}^{\text{DR}}(R, X, Y, Z, \zeta_0, \omega_0, \theta^*, \boldsymbol{u} = Y)\} = E\{Y\}.$$

Under model $\mathcal{M}_{\text{OR}}$, let ω^* denote the probability limit of estimation under model $P^*(r|Y=0, x, z; \omega)$. Then at true parameter values (ζ_0, θ_0),

$$
\begin{aligned}
&E\{\boldsymbol{G}^{\text{DR}}(R, X, Y, Z; \zeta_0, \omega^*, \theta_0, \boldsymbol{u})|X, Z\} \\
=&E\left\{ \frac{R}{\pi(\zeta_0, \omega^*)}\{\boldsymbol{u}(X,Y) - \rho(X,Z)\} + \rho(X,Z) \middle| X, Z \right\} \\
=&E\left\{ \frac{R\{1-\pi(\zeta_0, \omega^*)\}}{\pi(\zeta_0, \omega^*)}\{\boldsymbol{u}(X,Y) - \rho(X,Z)\} \middle| X, Z \right\} \\
&+ E\{\rho(X,Z) + R\{\boldsymbol{u}(X,Y) - \rho(X,Z)\}|X,Z\} \\
=&E\left\{ R\left\{ \mathrm{e}^{-\{\lambda(X,Z;\omega^*)+\eta(X,Y,Z;\zeta_0)\}} \right\} \{\boldsymbol{u}(X,Y) - \rho(X,Z)\} \middle| X, Z \right\} \\
&+ E\{\boldsymbol{u}(X,Y)|X,Z\} \\
=&E\{\boldsymbol{u}(X,Y)|X,Z\}. \qquad \text{(S4)}
\end{aligned}
$$

The estimating function for (4.12), under iterated expectations with respect to (X, Z) at $(\xi_0, \zeta_0, \theta_0)$, is

$$=E\left\{ [\boldsymbol{v}(X,Z) - E(\boldsymbol{v}(X,Z)|X)] \{E(\boldsymbol{u}(X,Y)|Z,X)\} \right\}$$

$$=E\left\{ [\boldsymbol{v}(X,Z) - E(\boldsymbol{v}(X,Z)|X)] \{E(\boldsymbol{u}(X,Y)|X)\} \right\} \qquad \text{by } \textbf{(IV.1)}$$

$$=E\left\{ [E(\boldsymbol{v}(X,Z)|X) - E(\boldsymbol{v}(X,Z)|X)] \{E(\boldsymbol{u}(X,Y)|X)\} \right\}$$

$$=\mathbf{0}.$$

In addition, under iterated expectations with respect to (X, Z) and with similar reasoning given in (S4),

$$E\left\{\boldsymbol{G}^{\text{DR}}(R, X, Y, Z, \zeta_0, \omega^*, \theta_0, \boldsymbol{u} = Y)\right\} = E\{Y\}.$$

The consistency and asymptotic normality of $\hat{\phi}^{\text{DR}}$ can be established under standard regularity conditions for GMM estimators (Newey and McFadden (1994)). A necessary condition is that the probability of observing the outcome is bounded away from zero (S1). Let $\boldsymbol{M}(\delta)$ represent the stacked vector of the following estimating functions: score functions for estimating ξ and θ, estimating function (4.7) for estimating ω, $\boldsymbol{U}^{\text{DR}}(\xi, \zeta, \theta, \omega)$ and $G(\phi, \zeta, \omega, \theta) = \{\boldsymbol{G}^{\text{DR}}(R, X, Y, Z; \zeta, \theta, \omega, \boldsymbol{u}^\dagger) - \phi\}$, where $\delta = (\omega, \zeta, \theta, \xi, \phi)$ and $\boldsymbol{u}^\dagger(X, Y) = Y$. Then under standard regularity conditions for M-estimation (Newey and McFadden (1994)), the asymptotic variance V_{DR} is given by the diagonal entry corresponding to ϕ of the following variance-covariance matrix

$$\left[E\left\{\frac{\partial \boldsymbol{M}(\delta)}{\partial \delta^T}\bigg|_{\delta_0}\right\}\right]^{-1} E\left\{\boldsymbol{M}(\delta_0)\boldsymbol{M}(\delta_0)^T\right\}\left[E\left\{\frac{\partial \boldsymbol{M}(\delta)}{\partial \delta^T}\right\}\bigg|_{\delta_0}\right]^{-1^T}, \quad \text{(S5)}$$

where $\delta_0 = (\omega_0, \zeta_0, \theta_0, \xi_0, \phi_0)$ is the probability limit of $\hat{\delta} = (\hat{\omega}, \hat{\zeta}, \hat{\theta}, \hat{\xi}, \hat{\phi})$. A consistent sandwich estimator for the above asymptotic variance can be constructed by evaluating unknown expectations as sample means at the estimated parameter value $\hat{\delta}$.

Bibliography

Newey, W. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics* (Edited by D. McFadden and R. Engler), volume 4. Elsevier Science.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–866.