Section 7.4: Optimal Propensity Score Estimation

Jae-Kwang Kim

Department of Statistics, Iowa State University

Improving the efficiency of PS estimator

- We want to improve the efficiency of the PS estimator by incorporating the auxiliary variable x_i observed throughout the sample.
- One can consider a class of estimating equations of the form

$$\frac{1}{n}\sum_{i=1}^{n}\delta_i\frac{1}{\hat{\pi}_i}\left\{U(\theta;x_i,y_i)-b(\theta;x_i)\right\}+\frac{1}{n}\sum_{i=1}^{n}b(\theta;x_i)=0 \qquad (1)$$

  where $b(\theta;x_i)$ is to be determined.
- We can write the solution to (1) as $\hat{\theta}_b$ as it depends on the particular choice of $b(\theta;x)$ function.
- Note that the solution $\hat{\theta}_b$ is consistent regardless of the choice of $b(\theta;x_i)$.
- We want to find an optimal choice $b^*(\theta;x_i)$ which minimizes the variance of $\hat{\theta}_b$.

Theorem 7.3 (Robins, Rotnitzky, and Zhao; 1994)

Theorem

Assume that the probability $Pr(\delta = 1 \mid x, y) = \pi(x)$ *does not depend on the value of* y. *Let* $\hat{\theta}_b$ *be the solution to (1) for given* $b(\theta; x_i)$. *Under some regularity conditions,* $\hat{\theta}_b$ *is consistent and its asymptotic variance satisfies*

$$V\left(\hat{\theta}_b\right) \geq n^{-1}\tau^{-1}\left[V\left\{E(U \mid X)\right\} + E\left\{\frac{1}{\pi(x)}V(U \mid X)\right\}\right](\tau^{-1})', \quad (2)$$

where $\tau = E(\partial U/\partial\theta')$ *and the equality holds when* $b^*(\theta; x_i) = E\left\{U(\theta; x_i, y_i) \mid x_i\right\}$.

In survey sampling, a similar result has been proved by Isaki and Fuller (1982).

Proof (Sketched)

- We first consider the case when the true response probability $\pi(x) = Pr(\delta = 1 \mid x)$ is known. Let

$$U_b(\theta) = \frac{1}{n}\sum_{i=1}^{n} b(\theta; x_i) + \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{\pi_i}\left\{U(\theta; x_i, y_i) - b(\theta; x_i)\right\}$$

  where $\pi_i = \pi(x_i)$.

- Since $E\{U_b(\theta_0)\} = E\left\{U(\theta_0; x, y)\right\} = 0$, we have, by Lemma 4.1, the solution $\hat{\theta}_b$ to $U_b(\theta) = 0$ is asymptotically unbiased for θ_0 and

$$V(\hat{\theta}_b) \cong \tau^{-1} V\{U_b(\theta_0)\}(\tau^{-1})'$$

  where $\tau = E\{\partial U_b(\theta_0)/\partial\theta'\} = E\{\partial U(\theta_0; x, y)/\partial\theta'\}$.

- Now, write

$$U_b(\theta) = U_{b^*}(\theta) + D_b(\theta) \tag{3}$$

where

$$\begin{aligned} U_{b^*}(\theta) &= \frac{1}{n}\sum_{i=1}^{n} b^*(\theta; x_i) + \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{\pi_i}\left\{U(\theta; x_i, y_i) - b^*(\theta; x_i)\right\} \\ D_b(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\delta_i}{\pi_i} - 1\right)\left\{b^*(\theta; x_i) - b(\theta; x_i)\right\} \end{aligned}$$

and $b^*(\theta; x_i) = E\left\{U(\theta; x_i, y_i) \mid x_i\right\}$.

- We have

$$V\{U_b(\theta)\} = V\{U_{b^*}(\theta)\} + V\{D_b(\theta)\} + 2Cov\left\{U_{b^*}(\theta), D_b(\theta)\right\}.$$

- Since

$$Cov\left\{U_{b^*}(\theta), D_b(\theta)\right\} = E\left\{n^{-2}\sum_{i=1}^{n}\left(\frac{1}{\pi_i}-1\right)(U_i - b_i^*)(b_i^* - b_i)\right\}$$

where $U_i = U(\theta; x_i, y_i)$, $b_i = b(\theta; x_i)$, and $b_i^* = b^*(\theta; x_i)$. Because $E(U_i \mid x_i) = b_i^*$, the above covariance term is equal to zero and

$$V\{U_b(\theta)\} \geq V\{U_{b^*}(\theta)\}.$$

- Since

$$
\begin{aligned}
V\{U_{b^*}(\theta)\} &= V\left\{\frac{1}{n}\sum_{i=1}^{n} U(\theta; x_i, y_i)\right\} \\
&\quad + E\left\{\frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{1}{\pi_i}-1\right)(U_i - b_i^*)^2\right\} \\
&= n^{-1}V(U) + n^{-1}E\left\{(\pi^{-1}-1)V(U \mid X)\right\} \\
&= n^{-1}V\{E(U \mid X)\} + n^{-1}E\left\{\pi^{-1}V(U \mid X)\right\},
\end{aligned}
$$

result (2) holds when the true response probability is used.

- To discuss the case with estimated response probability, let $\hat{\pi}_i = \pi(x_i; \hat{\phi})$ where $\hat{\phi}$ is estimated by solving $U_2(\phi) = 0$. Writing

$$
U_{1b}(\theta, \phi) = \frac{1}{n}\sum_{i=1}^{n} b(\theta; x_i) + \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi(x_i; \phi)}\left\{U(\theta; x_i, y_i) - b(\theta; x_i)\right\},
$$

the solution $(\hat{\theta}_b, \hat{\phi})$ to $U_{1b}(\theta, \phi) = 0$ and $U_2(\phi) = 0$ is consistent for (θ_0, ϕ_0).

- Using the standard Taylor linearization, it can be shown that

$$V(\hat{\theta}_b) \cong \tau^{-1} V\{U_{1b}(\theta_0, \phi_0) - CU_2(\phi_0)\}(\tau^{-1})' \tag{4}$$

where $C = (\partial U_{1b}/\partial \phi')(\partial U_2/\partial \phi')^{-1}$. Now, similarly to (3), we can write

$$U_{1b}(\theta, \phi) - CU_2(\phi) = U_{b^*}(\theta) + D_{2b}(\theta, \phi)$$

where $D_{2b}(\theta, \phi) = D_b(\theta) - CU_2(\phi)$.

- Because $U_2(\phi)$ does not depend on y, we can show that

$$Cov\,(U_{b^*}, D_{2b}) = E\{n^{-2} \sum_{i=1}^{n} (\pi_i^{-1} - 1)(U_i - b_i^*)(b_i^* - b_i - C\mathbf{h}_i)\} = 0$$

and result (2) follows.

Example 7.4

- Consider the sample from a linear regression model

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + e_i \tag{5}$$

  where e_i are independent with $E(e_i \mid \mathbf{x}_i) = 0$. Assume that $\mathbf{x}_i$ are available from the full sample and y_i are observed only when $\delta_i = 1$. The response propensity model follows from the logistic regression model with $\text{logit}(\pi_i) = \mathbf{x}_i'\phi$. We are interested in estimating $\theta = E(Y)$ from the partially observed data.

- To construct the optimal estimator that achieves the minimum variance in (2), we can use $U_i(\theta) = y_i - \theta$ and $b_i^*(\theta) = \mathbf{x}_i'\boldsymbol{\beta} - \theta$. Thus, the optimal estimator using $\hat{b}_i^*(\theta) = \mathbf{x}_i'\hat{\boldsymbol{\beta}} - \theta$ in (1) is given by

$$\hat{\theta}_{opt}(\hat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i}y_i + \frac{1}{n}\left(\sum_{i=1}^{n}\mathbf{x}_i - \sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i}\mathbf{x}_i\right)'\hat{\boldsymbol{\beta}} \tag{6}$$

  where $\hat{\boldsymbol{\beta}}$ is any estimator of $\boldsymbol{\beta}$ satisfying $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O_p(1)$, where $X_n = O_p(1)$ denotes that X_n is bounded in probability.

Example 7.4 (Cont'd)

- Note that the choice of $\hat{\boldsymbol{\beta}}$ does not play any leading role in the asymptotic variance of $\hat{\theta}_{opt}(\hat{\boldsymbol{\beta}})$. This is because

$$\hat{\theta}_{opt}(\hat{\boldsymbol{\beta}}) \cong \hat{\theta}_{opt}(\boldsymbol{\beta}_0) + E\left\{\frac{\partial}{\partial \boldsymbol{\beta}}\hat{\theta}_{opt}(\boldsymbol{\beta}_0)\right\}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \tag{7}$$

and, under the correct response model,

$$E\left\{\frac{\partial}{\partial \boldsymbol{\beta}}\hat{\theta}_{opt}(\boldsymbol{\beta}_0)\right\} = E\left\{\frac{1}{n}\left(\sum_{i=1}^{n}\mathbf{x}_i - \sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i}\mathbf{x}_i\right)\right\} \cong \mathbf{0}$$

and so the second term of (7) becomes negligible. Furthermore, it can be shown that the choice of $\hat{\phi}$ in $\hat{\pi}_i = \pi_i(\hat{\phi})$ does not matter as long as the regression model holds.

Remark

- In Example 7.4, optimal estimator using auxiliary information is considered, under the (outcome) regression model

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + e_i$$

  with $E(e_i \mid \mathbf{x}_i) = 0$.
- We now want to find the optimal estimator (using auxiliary information) without relying on the outcome regression model.
- Note that $\hat{\theta}_{PSA} = n^{-1}\sum_{i=1}^{n} \delta_i y_i/\hat{\pi}_i$ applied to $y_i = x_i$ does not necessarily lead to $\bar{x}_n = n^{-1}\sum_{i=1}^{n} x_i$.
- That is, we have two estimators of $E(X)$, $\hat{\bar{x}}_{PSA}$ and $\bar{x}_n$.
- How to incorporate the extra information without relying on the regression model ?

Introduction

- θ_0: p-dimensional parameter
- $U(\theta; Z) = 0$: a system of estimating equations satisfying $E\{U(\theta_0; Z)\} = 0$.
- Let m be the rank of $U(\theta; Z)$.
  1. $m = p$ (just identified): Unique solution
  2. $m < p$ (under-identified): Many solutions, no unique solution.
  3. $m > p$ (over-identified): No exact solution
- How to handle the over-identified case?
  1. Use GMM (Generalized Method of Moments)
  2. Use EL (Empirical Likelihood)

GMM (Generalized method of moment) estimation

- Let W be a $m \times m$ symmetric matrix. Define

$$Q_W(\theta) = \{U(\theta; Z)\}' W U(\theta; Z).$$

- GMM estimator:

$$\hat{\theta}_W = \arg\min Q_W(\theta)$$

- Note that the solution $\hat{\theta}_W$ is obtained by solving

$$U_W(\theta) \equiv \{\dot{U}(\theta; Z)\}' W U(\theta; Z) = 0$$

where $\dot{U}(\theta; z) = \partial U(\theta; z)/\partial \theta$.

GMM estimation (Cont'd)

- Thus, the asymptotic variance of $\hat{\theta}_W$ is

$$
\begin{aligned}
V\left(\hat{\theta}_W\right) &\cong \left\{E\left(\frac{\partial}{\partial\theta'}U_W(\theta_0)\right)\right\}^{-1} V\{U_W(\theta_0)\}\left\{E\left(\frac{\partial}{\partial\theta'}U_W(\theta_0)\right)'\right\} \\
&= \left\{\tau' W\tau\right\}^{-1}\tau' WV(U)W\tau\left\{\tau' W\tau\right\}^{-1},
\end{aligned}
$$

where $\tau = E(\partial U/\partial\theta)$. The asymptotic variance is minimized at

$$
W^* = \{V(U)\}^{-1}.
$$

- Minimize $\tau' WVW\tau$ subject to $\tau' W\tau$=Constant $\Rightarrow$ $W^* \propto V^{-1}$.

GMM estimation (Cont'd)

- Optimal GMM estimator:

$$\hat{\theta}^* = \arg\min Q_{W^*}(\theta)$$

where

$$\begin{aligned} Q_{W^*}(\theta) &= \{U(\theta; Z)\}' W^* U(\theta; Z) \\ &= \{U(\theta; Z)\}' \left[V\{U(\theta)\}\right]^{-1} U(\theta; Z) \end{aligned}$$

- The asymptotic variance of the optimal GMM estimator is $\left[\tau' \{V(U)\}^{-1} \tau\right]^{-1}$.

Empirical likelihood approach

- Instead of using GMM, empirical likelihood (EL) method can be also used to handle over-identified estimating equation problem.

- EL approach: Find θ that maximizes the profile empirical likelihood function of θ:

$$L(\theta) = \max \left\{ \prod_{i=1}^{n} p_i ; p_i > 0, \sum_{i=1}^{n} p_i = 1, \text{ and } \sum_{i=1}^{n} p_i U(\theta; z_i) = 0 \right\}. \tag{8}$$

Qin and Lawless (1994)

Theorem

Under some regularity conditions, the maximizer $\hat{\theta}$ of $L(\theta)$ in (8) satisfies

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) \xrightarrow{\mathcal{L}} N\left(0, V\right),$$

where

$$V=\left[E\left(\frac{\partial}{\partial\theta}U\right)'\left\{V\left(U\right)\right\}^{-1}E\left(\frac{\partial}{\partial\theta}U\right)\right]^{-1}.$$

Note: The maximum EL estimator is asymptotically equivalent to the optimal GMM estimator.

Lemma 7.4

Lemma

Assume that $\hat{X}_1$ *and* $\hat{X}_2$ *are two unbiased estimators of* μ_x *and* $\hat{Y}$ *is an unbiased estimator of* μ_y. *Let*

$$Q = \begin{pmatrix} \hat{X}_1 - \mu_x \\ \hat{X}_2 - \mu_x \\ \hat{Y} - \mu_y \end{pmatrix}' \begin{pmatrix} V(\hat{X}_1) & C(\hat{X}_1, \hat{X}_2) & C(\hat{X}_1, \hat{Y}) \\ C(\hat{X}_1, \hat{X}_2) & V(\hat{X}_2) & C(\hat{X}_2, \hat{Y}) \\ C(\hat{X}_1, \hat{Y}) & C(\hat{X}_2, \hat{Y}) & V(\hat{Y}) \end{pmatrix}^{-1} \begin{pmatrix} \hat{X}_1 - \mu_x \\ \hat{X}_2 - \mu_x \\ \hat{Y} - \mu_y \end{pmatrix}. \tag{9}$$

The optimal estimator of μ_x *that minimizes* Q *in (9) is*

$$\hat{\mu}_x^* = \alpha^* \hat{X}_1 + (1 - \alpha^*) \hat{X}_2 \tag{10}$$

where

$$\alpha^* = \frac{V(\hat{X}_2) - C(\hat{X}_1, \hat{X}_2)}{V(\hat{X}_1) + V(\hat{X}_2) - 2C(\hat{X}_1, \hat{X}_2)}.$$

Lemma (Cont'd)

Also, the optimal estimator of μ_y is

$$\hat{\mu}_y^* = \hat{Y} + B_1 \left(\hat{\mu}_x^* - \hat{X}_1 \right) + B_2 \left(\hat{\mu}_x^* - \hat{X}_2 \right), \tag{11}$$

where $\hat{\mu}_x^*$ is defined in (10) and

$$\begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = \begin{pmatrix} V(\hat{X}_1) & C(\hat{X}_1, \hat{X}_2) \\ C(\hat{X}_1, \hat{X}_2) & V(\hat{X}_2) \end{pmatrix}^{-1} \begin{pmatrix} C(\hat{X}_1, \hat{Y}) \\ C(\hat{X}_2, \hat{Y}) \end{pmatrix}.$$

Proof

Using the inverse of the partitioned matrix, we can write

$$Q(\mu_x, \mu_y) = Q_1(\mu_x) + Q_2(\mu_y \mid \mu_x)$$

where

$$\begin{aligned} Q_1 &= \begin{pmatrix} \hat{X}_1 - \mu_x \\ \hat{X}_2 - \mu_x \end{pmatrix}' \begin{pmatrix} V(\hat{X}_1) & C(\hat{X}_1, \hat{X}_2) \\ C(\hat{X}_1, \hat{X}_2) & V(\hat{X}_2) \end{pmatrix}^{-1} \begin{pmatrix} \hat{X}_1 - \mu_x \\ \hat{X}_2 - \mu_x \end{pmatrix}, \\ Q_2 &= \left\{ \hat{Y} - E(\hat{Y} \mid \hat{X}_1, \hat{X}_2) \right\}' V_{ee}^{-1} \left\{ \hat{Y} - E(\hat{Y} \mid \hat{X}_1, \hat{X}_2) \right\}, \end{aligned}$$

and

$$\begin{aligned} E(\hat{Y} \mid \hat{X}_1, \hat{X}_2) &= \mu_y + B_1(\hat{X}_1 - \mu_x) + B_2(\hat{X}_2 - \mu_x), \\ V_{ee} &= V(\hat{Y}) - (B_1, B_2)\{V(\hat{X}_1, \hat{X}_2)\}^{-1}(B_1, B_2)'. \end{aligned}$$

Proof

- Minimizing Q_1 with respect to μ_x gives $\hat{\mu}_x^*$ in (10).
- Minimizing Q_2 with respect to μ_y for given $\hat{\mu}_x^*$ gives $\hat{\mu}_y^*$ in (11).

Remark

- The optimal estimator of μ_y takes the form of the regression estimator with $\hat{\mu}_x^*$ as the control.
- Using (10), we can also express

$$\hat{\mu}_y^* = \hat{Y} - C\left(\hat{Y}, \hat{X}_2 - \hat{X}_1\right)\left\{V\left(\hat{X}_2 - \hat{X}_1\right)\right\}^{-1}(\hat{X}_2 - \hat{X}_1).$$

Applying GMM under missing data

- Under the missing data setup where $\mathbf{x}_i$ is always observed and y_i is subject to missingness, if we know π_i, then we can use $\hat{X}_1 = n^{-1}\sum_{i=1}^n \mathbf{x}_i = \hat{X}_n$, $\hat{X}_2 = n^{-1}\sum_{i=1}^n \delta_i \mathbf{x}_i/\pi_i = \hat{X}_W$, and $\hat{Y} = n^{-1}\sum_{i=1}^n \delta_i y_i/\pi_i = \hat{Y}_W$.
- In this case, we can obtain $\hat{\mu}_x^* = \bar{X}_1$ and the optimal estimator of μ_y reduces to

$$\begin{aligned}\hat{\mu}_y^* &= \hat{Y} + C\left(\hat{Y}, \hat{X}_2 - \hat{X}_1\right)\left\{V\left(\hat{X}_2 - \hat{X}_1\right)\right\}^{-1}(\hat{X}_1 - \hat{X}_2) \\ &= \hat{Y}_W + (\hat{X}_n - \hat{X}_W)' B^*\end{aligned}$$

where

$$B^* = E\left(\sum_{i=1}^n \frac{1-\pi_i}{\pi_i}\mathbf{x}_i\mathbf{x}_i'\right)^{-1} E\left(\sum_{i=1}^n \frac{1-\pi_i}{\pi_i}\mathbf{x}_i y_i\right).$$

Optimal PS estimation

GMM approach

- Let $\theta = (\mu_x, \mu_y)$. We have three estimators for two parameters.
- Find θ that minimizes

$$Q_{PS}(\theta) = \begin{pmatrix} \bar{x}_n - \mu_x \\ \hat{\theta}_{x,PS} - \mu_x \\ \hat{\theta}_{y,PS} - \mu_y \end{pmatrix}' \left\{ \hat{V} \begin{pmatrix} \bar{x}_n \\ \hat{\theta}_{x,PS} \\ \hat{\theta}_{y,PS} \end{pmatrix} \right\}^{-1} \begin{pmatrix} \bar{x}_n - \mu_x \\ \hat{\theta}_{x,PS} - \mu_x \\ \hat{\theta}_{y,PS} - \mu_y \end{pmatrix} \quad (12)$$

  where $\hat{\theta}_{PS} = \hat{\theta}_{PS}(\hat{\phi})$ is the propensity score estimator using $\hat{\pi}_i$.
- Computation for $\hat{V}$ is somewhat cumbersome.

Optimal PS estimation (Cont'd)

Alternative GLS (or GMM) approach

- Find (θ, ϕ) that minimizes

$$Q^* = \begin{pmatrix} \bar{x}_n - \mu_x \\ \hat{\theta}_{x,PS}(\phi) - \mu_x \\ \hat{\theta}_{y,PS}(\phi) - \mu_y \\ S(\phi) \end{pmatrix}' \left\{ \hat{V} \begin{pmatrix} \bar{x}_n \\ \hat{\theta}_{x,PS}(\phi) \\ \hat{\theta}_{y,PS}(\phi) \\ S(\phi) \end{pmatrix} \right\}^{-1} \begin{pmatrix} \bar{x}_n - \mu_x \\ \hat{\theta}_{x,PS}(\phi) - \mu_x \\ \hat{\theta}_{y,PS}(\phi) - \mu_y \\ S(\phi) \end{pmatrix}$$

- Computation for $\hat{V}$ is easier since we can treat ϕ as if known.
- Let $Q^*(\theta, \phi)$ be the above objective function. It can be shown that $Q^*(\theta, \hat{\phi}) = Q_{PS}(\theta)$ in (12), which implies that minimizing $Q^*(\theta, \hat{\phi})$ is equivalent to minimizing $Q_{PS}(\theta)$.

Optimal PS estimation (Cont'd)

Justification for the equivalence

- May write

$$
\begin{aligned}
Q^*(\theta,\phi) &= \begin{pmatrix} \hat{U}_{PS}(\theta,\phi) \\ S(\phi) \end{pmatrix}' \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}^{-1} \begin{pmatrix} \hat{U}_{PS}(\theta,\phi) \\ S(\phi) \end{pmatrix} \\
&= Q_1(\theta \mid \phi) + Q_2(\phi)
\end{aligned}
$$

where

$$
\begin{aligned}
Q_1(\theta \mid \phi) &= \left(\hat{U}_{PS} - V_{12}V_{22}^{-1}S\right)' \left\{V\left(U_{PS} \mid S^{\perp}\right)\right\}^{-1} \left(\hat{U}_{PS} - V_{12}V_{22}^{-1}S\right) \\
Q_2(\phi) &= S(\phi)' \left\{\hat{V}\left(S\right)\right\}^{-1} S(\phi)
\end{aligned}
$$

- For the MLE $\hat{\phi}$, we have $Q_2(\hat{\phi}) = 0$ and $Q_1(\theta \mid \hat{\phi}) = Q_{PS}(\theta)$.

Example 7.5

- Response model

$$\pi_i(\phi^*) = \frac{\exp(\phi_0^* + \phi_1^* x_i)}{1 + \exp(\phi_0^* + \phi_1^* x_i)}$$

- Three direct PS estimators of $(1, \mu_x, \mu_y)$:

$$(\hat{\theta}_{1,PS}, \hat{\theta}_{x,PS}, \hat{\theta}_{y,PS}) = n^{-1} \sum_{i=1}^{n} \delta_i \hat{\pi}_i^{-1} (1, x_i, y_i).$$

- $\bar{x}_n = n^{-1} \sum_{i=1}^{n} x_i$ available.
- What is the optimal estimator of μ_y ?

Example 7.5 (Cont'd)

- Minimize

$$\begin{pmatrix} \bar{x}_n - \mu_x \\ \hat{\theta}_{1,PS}(\phi) - 1 \\ \hat{\theta}_{x,PS}(\phi) - \mu_x \\ \hat{\theta}_{y,PS}(\phi) - \mu_y \\ S(\phi) \end{pmatrix}' \left\{ \hat{V} \begin{pmatrix} \bar{x}_n \\ \hat{\theta}_{1,PS}(\phi) \\ \hat{\theta}_{x,PS}(\phi) \\ \hat{\theta}_{y,PS}(\phi) \\ S(\phi) \end{pmatrix} \right\}^{-1} \begin{pmatrix} \bar{x}_n - \mu_x \\ \hat{\theta}_{1,PS}(\phi) - 1 \\ \hat{\theta}_{x,PS}(\phi) - \mu_x \\ \hat{\theta}_{y,PS}(\phi) - \mu_y \\ S(\phi) \end{pmatrix}$$

with respect to (μ_x, μ_y, ϕ), where

$$S(\phi) = \sum_{i=1}^{n} \left(\frac{\delta_i}{\pi_i(\phi)} - 1 \right) \mathbf{h}_i(\phi) = 0$$

with $\mathbf{h}_i(\phi) = \pi_i(\phi)(1, x_i)'$.

Example 7.5 (Cont'd)

- Equivalently, minimize

$$\begin{pmatrix} \hat{\theta}_{y,PS}(\phi) - \mu_y \\ \hat{\theta}_{1,PS}(\phi) - 1 \\ \hat{\theta}_{x,PS}(\phi) - \bar{x}_n \\ S(\phi) \end{pmatrix}' \left\{ \hat{V} \begin{pmatrix} \hat{\theta}_{y,PS}(\phi) \\ \hat{\theta}_{1,PS}(\phi) \\ \hat{\theta}_{x,PS}(\phi) - \bar{x}_n \\ S(\phi) \end{pmatrix} \right\}^{-1} \begin{pmatrix} \hat{\theta}_{y,PS}(\phi) - \mu_y \\ \hat{\theta}_{1,PS}(\phi) - 1 \\ \hat{\theta}_{x,PS}(\phi) - \bar{x}_n \\ S(\phi) \end{pmatrix}$$

with respect to (μ_y, ϕ), since the optimal estimator of θ_x is $\bar{x}_n$.

Example 7.5 (Cont'd)

- The solution can be written as

$$\hat{\mu}_{y,opt} = \hat{\theta}_{y,PS} + \left(1 - \hat{\theta}_{1,PS}\right)\hat{B}_0 + \left(\bar{x}_n - \hat{\theta}_{1,PS}\right)\hat{B}_1 + \left\{0 - S(\hat{\phi})\right\}\hat{C}$$

where

$$\begin{pmatrix} \hat{B}_0 \\ \hat{B}_1 \\ \hat{C} \end{pmatrix} = \left\{\sum_{i=1}^n \delta_i b_i \begin{pmatrix} 1 \\ x_i \\ \mathbf{h}_i \end{pmatrix} \begin{pmatrix} 1 \\ x_i \\ \mathbf{h}_i \end{pmatrix}'\right\}^{-1} \sum_{i=1}^n \delta_i b_i \begin{pmatrix} 1 \\ x_i \\ \mathbf{h}_i \end{pmatrix} y_i$$

and $b_i = \hat{\pi}_i^{-2}(1 - \hat{\pi}_i)$.

- Note that the last term $\{0 - S(\hat{\phi})\}\hat{C}$, which is equal to zero, does not contribute to the point estimation. But, it is used for variance estimation.

Example 7.5 (Cont'd)

- That is, for variance estimation, we simply express

$$\hat{\mu}_{y,opt} = n^{-1} \sum_{i=1}^{n} \hat{\eta}_i$$

where

$$\hat{\eta}_i = \hat{B}_0 + x_i \hat{B}_1 + \mathbf{h}_i' \hat{C} + \frac{\delta_i}{\hat{\pi}_i} \left(y_i - \hat{B}_0 - x_i \hat{B}_1 - \mathbf{h}_i' \hat{C} \right)$$

and apply the standard variance formula to $\hat{\eta}_i$.

Example 7.5 (Cont'd)

- The optimal estimator is linear in y. That is, we can write

$$\hat{\mu}_{y,opt} = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i}g_i y_i = \sum_{\delta_i=1} w_i y_i$$

  where g_i satisfies

$$\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i}g_i(1, x_i, \mathbf{h}_i') = \sum_{i=1}^{n}(1, x_i, \mathbf{h}_i').$$

- Thus, it is doubly robust under the outcome model $E(y \mid x) = \beta_0 + \beta_1 x$ in the sense that $\hat{\mu}_{y,opt}$ is unbiased when either the response model or the outcome model holds.