Comparison of Asymptotic Variance: MIPW vs IPW

Jaehyuk Jang

October 2024

1 Introduction

Proposition 1 (The Equivalent Estimating Equation for MIPW). *Let* $(Y_i, Y_i^*, \mathbf{X}_i, \mathbf{X}_i^*, Z_i) \overset{i.i.d}{\sim} F, i = 1, \ldots, N$. *Define the true parameter* $\theta_0 = (\beta_0, \pi_0, \mu_0(1), \mu_0(0)) \in \Theta \subset \mathbb{R}^{d+3}$ *as* $E_F(\psi^*(\theta_0)) = 0$. *Then,* ψ^{**} *is an equivalent estimating equation of* ψ^* *where* ψ^{**} *is*

$$\psi^{**}(\theta) = \begin{pmatrix} \{\frac{1-\delta}{e^*(\mathbf{X};\beta)} Z + (\frac{\delta\pi}{(1-\pi)e^*(\mathbf{X};\beta)} - \frac{1}{1-e^*(\mathbf{X};\beta)})(1-Z)\}\nabla_\beta e^*(\mathbf{X};\beta) \\ Z - \pi \\ ZY - Z\mu(1) \\ \frac{e(\mathbf{X};\beta)}{1-e(\mathbf{X};\beta)}(1-Z)Y - \frac{e(\mathbf{X};\beta)}{1-e(\mathbf{X};\beta)}(1-Z)\mu(0) \end{pmatrix}$$

This is the proposition that let us to calculate the MIPW estimator through only observed variance. By the M-estimation theory, $\sqrt{n}^{-1}(\hat{\tau} - \tau) \xrightarrow{d} N(0, V_{MIPW})$ where

$$V_{MIPW} = V(\psi^{**}) = a^T A(\psi^{**}) B(\psi^{**}) A(\psi^{**})^T a \tag{1}$$

$$A(\psi^{**}) = \left(E\left[-\frac{\partial}{\partial\theta}\psi^{**}\right]\right)^{-1} \tag{2}$$

$$B(\psi^{**}) = E\left[\psi^{**}(\psi^{**})^T\right] \tag{3}$$

$$a = \left(\mathbf{0}_d^T, 0, 1, -1\right)^T \tag{4}$$

We want to prove that $V_{MIPW} < V_{IPW}$ when δ is “bounded” away from 1. However, as we can observe from the simulation table 1 below, it is more likely that there is some range of δ such inequality holds. The conducted simulation proceeded as so.

1. Set $\delta = 0.1$. Generate 200 datasets with size $n = 500$ from the “strong overlap” condition.

2. Calculate the sandwich variance estimator of IPW and MIPW for each datasets i.e. $\hat{V}_{IPW}^1, \ldots, \hat{V}_{IPW}^{200}$ and $\hat{V}_{MIPW}^1, \ldots, \hat{V}_{MIPW}^{200}$.

3. Average the results and compare: $\hat{V}_{IPW} = \sum_r \hat{V}^r_{IPW}$ vs $\hat{V}_{MIPW} = \sum_r \hat{V}^r_{MIPW}$.
4. Repeat $1, 2, 3$ for $\delta = 0.6$ and 0.9.
5. Repeat the whole process for $n = 1000, 5000, 10000, 100000$.

δ	N	500	10^3	5×10^3	10^4	10^5
$\delta = 0.1$	$\hat{V}_{IPW}$	0.0487	0.0266	0.0051	0.0025	0.0003
	$\hat{V}_{MIPW}$	0.0385	0.0217	0.0042	0.0021	0.0002
$\delta = 0.6$	$\hat{V}_{IPW}$	0.0485	0.0237	0.0050	0.0025	0.0003
	$\hat{V}_{MIPW}$	0.0218	0.0106	0.0023	0.0011	0.0001
$\delta = 0.9$	$\hat{V}_{IPW}$	0.0509	0.0261	0.0050	0.0025	0.0003
	$\hat{V}_{MIPW}$	8.7025	1.9699	0.0108	0.0051	0.0005

Table 1: Simulation Result

This simulation result implies that $V_{MIPW} < V_{IPW}$ does not hold for all $\delta \in (0, 1)$ but hopefully on certain $(0, \tilde{\delta})$ for some constant $\tilde{\delta} < 1$.

2 Idea

2.1 Notations

Recall the estimating equation for the IPW estimator.

$$\psi(\theta) = \psi(\theta; Y, \mathbf{X}, Z) = \begin{pmatrix} \frac{Z - e(\mathbf{X};\beta)}{e(\mathbf{X};\beta)(1-e(\mathbf{X};\beta))} \nabla_\beta e(\mathbf{X}; \beta) \\ ZY - Z\mu(1) \\ \frac{e(\mathbf{X};\beta)}{1-e(\mathbf{X};\beta)}(1-Z)Y - \frac{e(\mathbf{X};\beta)}{1-e(\mathbf{X};\beta)}(1-Z)\mu(0) \end{pmatrix} \tag{5}$$

Denote

$$\begin{aligned} e &:= e(\mathbf{X}; \beta), \quad e^* := e^*(\mathbf{X}; \beta) \\ \dot{e} &:= \frac{\partial}{\partial \beta} e(\mathbf{X}; \beta) = \nabla_\beta e(\mathbf{X}; \beta) \\ \dot{e}^* &:= \frac{\partial}{\partial \beta} e^*(\mathbf{X}; \beta) = \nabla_\beta e^*(\mathbf{X}; \beta) \end{aligned}$$

Note from Lemma 1 in the main paper,

$$e^* = \frac{\pi\delta + (1 - \pi - \delta)e}{1 - \pi + \pi\delta - \delta e} \tag{6}$$

$$\dot{e}^* = (1 - \delta) \left(\frac{1 - e^*}{1 - e} \right)^2 \dot{e} \tag{7}$$

Let

$$\mathbf{C}^* = \left[\begin{array}{c|c} c^* I_d & \mathbf{0}_{d\times 3} \\ \hline \mathbf{0}_{3\times d} & I_3 \end{array}\right] \cdots c^* = \frac{e^*(1-e)^2}{(1-\delta)^2(1-e^*)^2 e} \tag{8}$$

2.2 Discoveries

Some list of discoveries:

(1) $\psi = \mathbf{C}^* \psi^{**}$ for $0 \leq \delta < 1$

(2) $B(\psi) \succeq B(\psi^{**})$ equality iff $\delta = 0$

Proof of (1) Observe that ψ and ψ^{**} have the same form except for the first element. Let us examine the first elements of ψ and ψ^{**}, ψ_1 and ψ_1^{**}.

$$\begin{aligned} \psi_1 &= \left(\frac{Z}{e} - \frac{1-Z}{1-e}\right)\dot{e} \\ \psi_1^{**} &= \left(Z\frac{1-\delta}{e^*} + (1-Z)\frac{\delta\pi - (1-\pi+\delta\pi)e^*}{(1-\pi)e^*(1-e^*)}\right)\dot{e}^* \end{aligned}$$

Consider $Z = 1$. Then

$$\begin{aligned} \psi_1^{Z=1} &= \frac{\dot{e}}{e} \qquad (9) \\ &\overset{(7)}{=} \frac{e^*(1-e)^2}{(1-\delta)^2(1-e^*)^2 e}(1-\delta)\frac{\dot{e}^*}{e^*} \qquad (10) \\ &= \frac{e^*(1-e)^2}{(1-\delta)^2(1-e^*)^2 e}(\psi_1^{**})^{Z=1} \qquad (11) \end{aligned}$$

Denote $c^* = \frac{e^*(1-e)^2}{(1-\delta)^2(1-e^*)^2 e}$ and note that $c^* > 1$ a.s. We will use this later.

$$\begin{aligned} c^* &\overset{(6)}{=} \frac{1-\pi+\pi\delta-\delta e}{(1-\pi)(1-\delta)}\left(1 + \frac{\delta}{1-\delta}\frac{\pi}{1-\pi}\frac{1-e}{e}\right) > 1 \quad a.s \qquad (12) \\ &(\because e < 1 \quad a.s) \qquad (13) \end{aligned}$$

Now consider $Z = 0$.

$$
\begin{aligned}
\psi_1^{Z=0} &= -\frac{\dot{e}}{1-e} && (14)\\
&\overset{(7)}{=} -\frac{1-e}{(1-\delta)(1-e^*)^2}\dot{e}^* && (15)\\
&\overset{(6)}{=} -\frac{1-\pi}{(1-\pi-\delta+\delta e^*)(1-e^*)}\dot{e}^* && (16)\\
&\overset{(6)}{=} \frac{(1-\pi)e^*(1-e)}{(1-\delta)(1-\pi-\delta+\delta e^*)(1-e^*)e}\frac{\delta\pi-(1-\pi+\delta\pi)e^*}{(1-\pi)e^*(1-e^*)}\dot{e}^* && (17)\\
&= \frac{(1-\pi)e^*(1-e)}{(1-\delta)(1-\pi-\delta+\delta e^*)(1-e^*)e}(\psi_1^{**})^{Z=0} && (18)
\end{aligned}
$$

However, note that $\frac{(1-\pi)e^*(1-e)}{(1-\delta)(1-\pi-\delta+\delta e^*)(1-e^*)e} = c^*$. Hence, $\psi_1^{Z=0} = c^*(\psi_1^{**})^{Z=0}$. Since Z is a binary variable, $Z(1-Z)=0$. Therefore,

$$\psi_1 = c^* Z\psi_1^{**} + c^*(1-Z)\psi_1^{**} = c^*\psi_1^{**} \tag{19}$$

Then,

$$\psi = \mathbf{C}^*\psi^{**} \tag{20}$$

where

$$\mathbf{C}^* = \left[\begin{array}{c|c} c^* I_d & \mathbf{0}_{d\times 3} \\ \hline \mathbf{0}_{3\times d} & I_3 \end{array}\right] \tag{21}$$

Proof of (2)

$$B(\psi) \succeq B(\psi^{**}) \Leftarrow \psi\psi^T \succeq \psi^{**}(\psi^{**})^T \quad a.s \overset{(*)}{\Leftarrow} (1) \tag{22}$$

Proof of $(*)$

Lemma 1. *Let $x \in \mathbb{R}^d$ and diagonal $D \in \mathbb{R}^{d\times d}$ s.t. $D \succeq I$. If $y = Dx$ then $yy^T \succeq xx^T$ where equality holds iff $D = I$.*

Proof. WTS: $yy^T = Dxx^TD \succeq xx^T$ which is equivalent to $Dxx^TD - xx^T \succeq \mathbf{0}$

Note

$$
\begin{aligned}
&(D-I)xx^T(D-I) \succeq \mathbf{0}\\
\Leftrightarrow &Dxx^TD \succeq xx^TD + Dxx^T - xx^T
\end{aligned}
$$

Hence,

ETS: $xx^TD + Dxx^T - xx^T \succeq xx^T$ which is sufficed when $Dxx^T \succeq xx^T$ since Dxx^T is symmetric. In other words, we need to show that $(D-I)xx^T$ is a p.s.d.

This is directly straightforward since $rank((D-I)xx^T) \leq \min(rank(D-I), rank(xx^T)) = 1$ and the unique pair of eigenvalue and eigenvector of $(D-I)xx^T$ is $x^T(D-I)x$ and $(D-I)x$, respectively. ☐

If we can prove that $A(\psi)^T = DA(\psi^{**})^T$ for some diagonal $D \succeq I$ on $0 \leq \delta \leq \tilde{\delta}$, then $V_{MIPW} > V_{IPW}$ on the same range by Lemma 2.

Lemma 2. *Let $x \in \mathbb{R}^d$ and diagonal $D \in \mathbb{R}^{d\times d}$ s.t. $D \succeq I$. If $y = Dx$ then $x^TBx \leq y^TBy$ for a symmetric p.s.d B. The equality holds iff $D = I$.*

Proof. To show $x^TBx \leq x^TDBDx = y^TBy$, note that since B is a p.s.d, $\exists B^{1/2} : B = B^{1/2}(B^{1/2})^T$ and $B^{1/2}$ is also symmetric ($\because$ spectral decomposition of B). Set $v = (B^{1/2})^Tx$ then $x^TBx = v^Tv \leq v^TD^2v = x^TDBDx$ where $D^2 = DD$. The equality holds iff $D = I$. ☐

2.3 Simulation Study: Find D such that $A(\psi)^T = DA(\psi^{**})^T$

To see if $A(\psi)^T(A(\psi^{**}))^{-T}$ is a diagonal matrix on a certain range of δ, we conduct a simple simulation study.

1. Set $\delta = 0.1$. Generate 200 datasets with size $n = 500$ from the "strong overlap" condition.

2. Calculate the bread of the sandwich variance estimator of IPW and MIPW for each datasets i.e. $\hat{A}(\psi)^1, \ldots, \hat{A}(\psi)^{200}$ and $\hat{A}(\psi^{**})^1, \ldots, \hat{A}(\psi^{**})^{200}$.

3. Compute $\hat{D}^r = (\hat{A}(\psi)^r)^T(\hat{A}(\psi^{**})^r)^{-T}$

4. Extract the pair; minimum and the maximum value of the absolute diagonal and the non-diagonal elements from each $\hat{D}^r, r = 1, \ldots, 200$ i.e. $\Big(\min_i |\hat{D}^r_{ii}|, \max_i |\hat{D}^r_{ii}|\Big)$ and $\Big(\min_{i\neq j} |\hat{D}^r_{ij}|, \max_{i\neq j} |\hat{D}^r_{ij}|\Big)$.

5. Average the pairs

6. Repeat 1 to 5 for $\delta = 0.6$ and $\delta = 0.9$.

7. Repeat the whole process for $n = 1000, 5000, 10000, 100000$

We omit the result for $n = 1000, 5000, 10000$ for simplicity.

n		500		10^5	
		diag.	non-diag.	diag.	non-diag.
$\delta = 0.1$	max	1.00064	1.00064	1.00013	0.05981
	min	0.74169	0.00000	0.77822	0.00000
		diag.	non-diag.	diag.	non-diag.
$\delta = 0.6$	max	1.00063	0.24343	1.00026	0.07236
	min	0.16752	0.00000	0.17835	0.00000
		diag.	non-diag.	diag.	non-diag.
$\delta = 0.9$	max	1.00029	0.55657	1.00118	0.01334
	min	0.01385	0.00000	0.01676	0.00000

Table 2: Simulation Result

According to the simulation result, D is not a diagonal matrix. Thus, the conjecture made in this subsection is wrong.