Zou, H. (2006), "The adaptive lasso and its oracle properties," *Journal of the American Statistical Association*, 101, 1418–1429.

8 Appendix

In the appendix, we include the proofs for all the theorems and propositions.

8.1 Proof of Proposition 2

We first introduce some graphoid axioms (Pearl and Paz, 1985) we will use later:

$$\begin{aligned}
\text{Intersection:} \quad & D \perp\!\!\!\perp Y|W,Z; D \perp\!\!\!\perp W|Y,Z \Rightarrow D \perp\!\!\!\perp Y,W|Z, && (19)\\
\text{Contraction:} \quad & D \perp\!\!\!\perp Y|Z; D \perp\!\!\!\perp W|Y,Z \Rightarrow D \perp\!\!\!\perp Y,W|Z, && (20)\\
\text{Weak union:} \quad & D \perp\!\!\!\perp X \cup Y \mid Z \Rightarrow D \perp\!\!\!\perp X \mid Z \cup Y, && (21)\\
\text{Decomposition :} \quad & D \perp\!\!\!\perp X \cup Y \mid Z \Rightarrow D \perp\!\!\!\perp X \mid Z. && (22)
\end{aligned}$$

We first show that any superset of $pa(Y)$ in $pa(Y) \cup \mathcal{I}$ is sufficient to adjust for confounding:

$$D \perp\!\!\!\perp Y(d) \mid \boldsymbol{X}_{\mathcal{M}}, \tag{23}$$

where $X_{\mathcal{M}} = pa(Y) \cup X_{\mathcal{S}}$. We show this by contradiction. Assume D and $Y(d)$ are d-connected given $X_{\mathcal{M}}$. Due to Assumption 2, there is not a direct edge between D and $Y(d)$. Furthermore, D and $Y(d)$ are not ancestral to each other due to Assumptions 4 and 5. Then either the following scenarios occur:

- $Y(d) \leftarrow Q \cdots D$, where Q is a parent of $Y(d)$. Since $Q \in pa(Y) \subset X_{\mathcal{M}}$, this path is blocked by $X_{\mathcal{M}}$;
- $Y(d) \rightarrow Q \cdots D$. This is impossible since $X_i's$ are non-descendants of $Y(d)$.

We now show that a precision variable is independent of the treatment conditional on confounders (and other precision variables):

$$D \perp\!\!\!\perp X_{\tilde{\mathcal{P}}} \mid (pa(Y) \setminus X_{\tilde{\mathcal{P}}}), \tag{24}$$

where $X_{\tilde{\mathcal{P}}} = pa(Y) \setminus X_{\mathcal{S}}$.

To see this, note that if $i \in \tilde{\mathcal{P}} \subset \mathcal{P}$, by the definition of $\mathcal{P}$ we have D and X_i are d-separated given $pa(Y) \setminus X_i$, which implies $D \perp\!\!\!\perp X_i \mid pa(Y) \setminus X_i$. Without loss of generality, assume $\tilde{\mathcal{P}} = \{1, 2, 3, \ldots, d_0\}$. We then have

$$D \perp\!\!\!\perp X_1 \mid [X_2 \cup (pa(Y) \setminus X_{1,2})],$$
$$D \perp\!\!\!\perp X_2 \mid [X_1 \cup (pa(Y) \setminus X_{1,2})].$$

By the intersection property (19), we have $D \perp\!\!\!\perp X_{1,2} \mid (pa(Y) \setminus X_{1,2})$. Repeat this process $d_0 - 1$ time, we then have $D \perp\!\!\!\perp X_{\tilde{\mathcal{P}}} \mid pa(Y) \setminus X_{\tilde{\mathcal{P}}}$.

Combining (23) and (24), by the contraction property (20), we can show that adjusting for all the confounders and some precision variables are sufficient to control for confounding:

$$D \perp\!\!\!\perp Y(d) \mid (pa(Y) \setminus X_{\tilde{\mathcal{P}}}). \tag{25}$$

We now show that an instrument variable set is independent of a precision variable conditional on confounders and other precision variables:

$$X_{\tilde{\mathcal{I}}} \perp\!\!\!\perp X_j \mid (pa(Y) \setminus X_j), \tag{26}$$

where $\tilde{\mathcal{I}} \subset \mathcal{I}$, $j \in \mathcal{P}$.

We again show by contradiction. Assume there exists $X_j \in X_{\mathcal{P}}$ such that X_j and $X_{\tilde{\mathcal{I}}}$ are d-connected given $pa(Y) \setminus X_j$. By definition $X_{\mathcal{I}} \subset pa(D)$, there is a path $D \leftarrow X_{\tilde{\mathcal{I}}}$. Then D

and X_j are d-connected given $pa(Y) \setminus X_j$, which is a contradiction to the definition of $\mathcal{P}$.

We now show that a set of instruments is independent of a subset of precision variables conditional on confounders and other precision variables:

$$X_{\tilde{\mathcal{I}}} \perp\!\!\!\perp X_{\tilde{\mathcal{P}}} \mid pa(Y) \setminus X_{\tilde{\mathcal{P}}}, \tag{27}$$

where $\tilde{\mathcal{I}} \subset \mathcal{I}$. Again, without loss of generality, we assume $\tilde{\mathcal{P}} = \{1, 2, 3, \ldots, d_0\}$. We then have:

$$X_{\tilde{\mathcal{I}}} \perp\!\!\!\perp X_1 \mid [X_2 \cup (pa(Y) \setminus X_{1,2})]$$
$$X_{\tilde{\mathcal{I}}} \perp\!\!\!\perp X_2 \mid [X_1 \cup (pa(Y) \setminus X_{1,2})]$$

By the intersection property (19), we have $X_{\tilde{\mathcal{I}}} \perp\!\!\!\perp X_{1,2} \mid (pa(Y) \setminus X_{1,2})$. Repeat this process $d_0 - 1$ time, we then have $X_{\tilde{\mathcal{I}}} \perp\!\!\!\perp X_{\tilde{\mathcal{P}}} \mid (pa(Y) \setminus X_{\tilde{\mathcal{P}}})$.

Finally, we show

$$D \perp\!\!\!\perp Y(d) \mid X_{\mathcal{S}}. \tag{28}$$

We use the same argument we used when we proved (23), we can show that

$$D \perp\!\!\!\perp X_{\tilde{\mathcal{P}}} \mid \{X_{\mathcal{I}} \cup (pa(Y) \setminus X_{\tilde{\mathcal{P}}})\}. \tag{29}$$

This relationship holds as $pa(D) \subset \{X_{\mathcal{I}} \cup (pa(Y) \setminus X_{\tilde{\mathcal{P}}})\}$, which means $\{X_{\mathcal{I}} \cup (pa(Y) \setminus X_{\tilde{\mathcal{P}}})\}$ is a superset of $pa(D)$. By letting $\tilde{\mathcal{I}} = \mathcal{I}$ in (27), combining (27), (29) and contraction property (20), we have

$$(X_{\mathcal{I}} \cup D) \perp\!\!\!\perp X_{\tilde{\mathcal{P}}} \mid (pa(Y) \setminus X_{\tilde{\mathcal{P}}}). \tag{30}$$

Combining (30) and decomposition property (22), for $\tilde{\mathcal{I}} \subset \mathcal{I}$ we have

$$(X_{\tilde{\mathcal{I}}} \cup D) \perp\!\!\!\perp X_{\tilde{\mathcal{P}}} \mid (pa(Y) \setminus X_{\tilde{\mathcal{P}}}). \tag{31}$$

Again, we use weak union property (21), we have the the following result:

$$D \perp\!\!\!\perp X_{\tilde{\mathcal{P}}} \mid \{(pa(Y) \setminus X_{\tilde{\mathcal{P}}}) \cup X_{\tilde{\mathcal{I}}}\}.$$

We set $\tilde{\mathcal{I}} = \mathcal{S} \cap \mathcal{I} \subset \mathcal{I}$. We note that $(pa(Y) \setminus X_{\tilde{\mathcal{P}}}) \cup (X_{\mathcal{S}} \cap X_{\mathcal{I}}) = [pa(Y) \cap \{pa(Y) \cap X_{\mathcal{S}}^c\}^c] \cup (X_{\mathcal{S}} \cap X_{\mathcal{I}}) = (pa(Y) \cap X_{\mathcal{S}}) \cup (X_{\mathcal{S}} \cap X_{\mathcal{I}}) = X_{\mathcal{S}}$. The last equality holds because $\mathcal{S} \subset \mathcal{P} \cup \mathcal{I} \cup \mathcal{C}$. So we have

$$D \perp\!\!\!\perp X_{\tilde{\mathcal{P}}} \mid X_{\mathcal{S}}. \tag{32}$$

We note that $X_{\mathcal{S}} \subset X_{\mathcal{M}}$ and $X_{\mathcal{S}} \cup X_{\tilde{\mathcal{P}}} = X_{\mathcal{M}}$. Combining (32) and (23), by contraction property (20), we proved our result (28).

8.2 Proof of Proposition 3

We are going to prove this proposition by contradiction. We assume it is true, that there exists $\mathcal{C}'$ which is the subset of $\mathcal{C}$ such that

$$Y(d) \perp\!\!\!\perp D \mid \mathcal{C}',$$

where $d = 0, 1$. We know from the previous argument we know that $\mathcal{C} \setminus \mathcal{C}'$ is not empty. Without loss of generality, we assume $\mathcal{C} \setminus \mathcal{C}'$ contains at least one element X_1. We first show that for all $\mathcal{E} \subset pa(Y) \setminus \mathcal{C}'$, we have

$$X_{\mathcal{E}} \perp\!\!\!\perp D \mid \mathcal{C}' \tag{33}$$

We will show this by contradiction. We assume this is not true, that $X_{\mathcal{E}} \not\!\perp\!\!\!\perp D \mid \mathcal{C}'$. So we have $X_{\mathcal{E}}$ and D are d-connected given $\mathcal{C}'$. Since $\mathcal{E} \subset pa(Y) \setminus \mathcal{C}' \subset pa(Y)$, there is a direct edge $X_{\mathcal{E}} \to Y$, which means D and $Y(d)$ are d-connected given $\mathcal{C}'$. Given faithful assumption 6, we have $Y(d) \not\!\perp\!\!\!\perp D \mid \mathcal{C}'$ This is a contradiction to the given condition.

We set $\mathcal{E} = pa(Y) \setminus \mathcal{C}'$ and $\mathcal{M} = pa(Y)$ in (23). Given (33), (23) and contraction property (20), we have

$$X_{\mathcal{E}} \cup Y(d) \perp\!\!\!\perp D \mid \mathcal{C}'. \tag{34}$$

Combining decomposition property (22) and (34), we have

$$(X_{\mathcal{E}} \setminus X_1) \cup Y(d) \perp\!\!\!\perp D \mid \mathcal{C}'.$$

We then use weak union property (21)

$$Y(d) \perp\!\!\!\perp D \mid pa(Y) \setminus X_1, \tag{35}$$

where $d = 0, 1$. (35) suggests that X_1 and D are d-separated given $pa(Y) \setminus X_1$, or there will have a backdoor path $Y(d) \leftarrow X_1 \cdots D$ given $pa(Y) \setminus X_1$, which will violate (35) under faithful assumption 6. However, this result violates the fact that $X_1 \in \mathcal{C}$, which suggests X_1 and D are d-connected given $pa(Y) \setminus X_1$. This contradiction helps us finished the proof of this proposition.
$\square$

8.3 Proof of Proposition 4

Part A In this part, we will show $D \perp\!\!\!\perp Y(d) \mid X_{\mathcal{S}_I}$ for $d = 0, 1$. For $i \in \mathcal{I}' \subset \mathcal{I}$, we first show that an instrument variable is independent of outcome given a sufficient set $\mathcal{S}$:

$$X_i \perp\!\!\!\perp Y(d) \mid X_{\mathcal{S}} \text{ for } d = 0, 1. \tag{36}$$

We will show this by contradiction. Assume this is not true, that $Y(d)$ and X_i are dependent given $X_{\mathcal{S}}$. Given faithfulness assumption 6, we know that $Y(d)$ and X_i are d-connected given $X_{\mathcal{S}}$. Since $X_i \in pa(D)$, there is a direct path that $X_i \leftarrow D$, so there exists a back door path $D \leftarrow X_i \cdots Y(d)$ given $X_{\mathcal{S}}$, which means D and $Y(d)$ are d-connected given $X_{\mathcal{S}}$. Given faithfulness assumption 6, we know that $Y(d)$ and D are dependent given $X_{\mathcal{S}}$, which is a contradiction to the condition.

We are ready to show $D \perp\!\!\!\perp Y(d) \mid X_{\mathcal{S}_I}$ for $d = 0, 1$. Given (36), $Y(d) \perp\!\!\!\perp D \mid X_{\mathcal{S}}$, and faithful assumption 6, we know that $Y(d)$ and (X_i, D) are d-separated given $X_{\mathcal{S}}$. So we have

$$(X_{\mathcal{I}'} \cup D) \perp\!\!\!\perp Y(d) \mid X_{\mathcal{S}} \text{ for } d = 0, 1. \tag{37}$$

Combining (37) and weak union property (21), we have

$$D \perp\!\!\!\perp Y(d) \mid X_{\mathcal{S}_I} \text{ for } d = 0, 1. \qquad \square$$

Part B In this part, we will show $D \perp\!\!\!\perp Y(d) \mid X_{\mathcal{S}_P}$ for $d = 0, 1$. We first show that a precision variable is independent of treatment given sufficient set $\mathcal{S}$:

$$X_i \perp\!\!\!\perp D \mid X_{\mathcal{S}}, \tag{38}$$

where $i \in \mathcal{P}$. Again we show this by contradiction. Assume this is not true, that D and X_i are dependent given $X_{\mathcal{S}}$. Given faithfulness assumption 6, we know that D and X_i are d-connected given $X_{\mathcal{S}}$. Since $X_i \in pa(Y)$, there is a direct path that $X_i \leftarrow Y$, so there exists a back door path $Y \leftarrow X_i \cdots D$ given $X_{\mathcal{S}}$, which means D and $Y(d)$ are d-connected given $X_{\mathcal{S}}$. Given faithfulness assumption 6, we know that $Y(d)$ and D are dependent given $X_{\mathcal{S}}$, which is a contradiction to the condition.

Next, we show $D \perp\!\!\!\perp Y(d) \mid X_{\mathcal{S}_P}$. Given (38), $Y(d) \perp\!\!\!\perp D \mid X_{\mathcal{S}}$, and faithful assumption 6, we know that $(Y(d), X_i)$ and D are d-separated given $X_{\mathcal{S}}$. So we have

$$D \perp\!\!\!\perp (Y(d) \cup X_{\mathcal{P}'}) \mid X_{\mathcal{S}} \text{ for } d = 0, 1. \tag{39}$$

Combining (39) and weak union property (21), we have

$$D \perp\!\!\!\perp Y(d) \mid X_{\mathcal{S}_P} \text{ for } d = 0, 1. \qquad \square$$

8.4 Proof of Proposition 6

Assume $(\boldsymbol{X}|D = d) \sim N(\mathbf{u_d}, \boldsymbol{\Sigma})$, then $(\boldsymbol{X}_{\mathcal{S}}|D = d) \sim N(\widetilde{\mathbf{u}}_d, \widetilde{\boldsymbol{\Sigma}})$. So we have

$$\frac{P(D = 1|\boldsymbol{X}_{\mathcal{S}})}{P(D = 0|\boldsymbol{X}_{\mathcal{S}})} = \frac{P(\boldsymbol{X}_{\mathcal{S}}|D = 1)}{P(\boldsymbol{X}_{\mathcal{S}}|D = 0)} \frac{P(D = 1)}{P(D = 0)}.$$

Let $\frac{P(D=1)}{P(D=0)} = \exp(c)$, where c is some real constant. We have

$$\frac{P(D = 1|\boldsymbol{X}_{\mathcal{S}})}{P(D = 0|\boldsymbol{X}_{\mathcal{S}})} = \exp(c) \exp\{(\widetilde{\mathbf{u}}_1 - \widetilde{\mathbf{u}}_0)\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}_{\mathcal{S}} - \frac{1}{2}(\widetilde{\mathbf{u}}_1^T\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\mathbf{u}}_1 - \widetilde{\mathbf{u}}_0^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\mathbf{u}}_0)\}.$$

Let $\alpha_0 = c - \frac{1}{2}(\widetilde{\mathbf{u}}_1^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\mathbf{u}}_1 - \widetilde{\mathbf{u}}_0^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\mathbf{u}}_0)$, and $\boldsymbol{\alpha} = (\widetilde{\mathbf{u}}_1 - \widetilde{\mathbf{u}}_0)\widetilde{\boldsymbol{\Sigma}}^{-1}$. We have

$$\frac{P(D = 1|\boldsymbol{X}_{\mathcal{S}})}{1 - P(D = 1|\boldsymbol{X}_{\mathcal{S}})} = \exp(\alpha_0 + \boldsymbol{X}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\alpha}).$$

Then we finish our proof. $\square$

8.5 Proof of Theorem 1

Without loss of generality, We assume $\mathcal{S} = \{1, 2, \ldots, d-1\}$ and $\mathcal{S}_I = \{1, 2, \ldots, p_0 - 1\}$, i.e. we add precision variables $X_i, i = d, \ldots, p_0 - 1$ into the set $\mathcal{S}$. We prove this result using standard M-estimation theories. In general, an M-estimator $\hat{\boldsymbol{\theta}}$ satisfies the following estimating equations

$$\sum_{i=1}^{n} \boldsymbol{\phi}(\boldsymbol{Y}_i, \hat{\boldsymbol{\theta}}) = 0.$$

Denote $\boldsymbol{\theta}_0$ the solution of vector function $E\{\boldsymbol{\phi}(\boldsymbol{Y}, \boldsymbol{\theta})\} = 0$. Stefanski and Boos (2002) showed that an M-estimator is asymptotically normally distributed with $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N\{0, V(\boldsymbol{\theta}_0)\}$, where $V(\boldsymbol{\theta}_0) = A(\boldsymbol{\theta}_0)^{-1}B(\boldsymbol{\theta}_0)\{A(\boldsymbol{\theta}_0)^{-1}\}^{\mathrm{T}}$, $A(\boldsymbol{\theta}_0) = E\{-\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\boldsymbol{\phi}(\boldsymbol{Y}, \boldsymbol{\theta}) \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\}$, $B(\boldsymbol{\theta}_0) = E\{\boldsymbol{\phi}(\boldsymbol{Y}, \boldsymbol{\theta})\boldsymbol{\phi}(\boldsymbol{Y}, \boldsymbol{\theta})^{\mathrm{T}} \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\}$.

Part A For the estimator (2), the corresponding estimating equations are $\boldsymbol{\phi}(Y, D, \boldsymbol{X}_{\mathcal{S}}; \boldsymbol{\theta}_{\mathcal{S}})$ when we try to estimate $e(\boldsymbol{X}_{\mathcal{S}}; \boldsymbol{\beta})$, we write

$$\begin{cases} \phi_0 = -\Delta_{\mathcal{S}} + \lambda \frac{YD}{e} - \kappa \frac{Y(1-D)}{1-e} \\ \phi_i = (\frac{D}{e} - \frac{1-D}{1-e})\partial e/\beta_i \\ \phi_{d+1} = -\lambda \frac{D}{e} + 1 \\ \phi_{d+2} = -\kappa \frac{1-D}{1-e} + 1, \end{cases} \tag{40}$$

where $1 \le i \le d$, β_1 is the intercept, β_{i+1} be the coefficient of i'th component of $\boldsymbol{X}_{\mathcal{S}}$. We write these equations to estimate M-estimator $\boldsymbol{\theta}_{\mathcal{S}} = (\Delta_{\mathcal{S}}, \boldsymbol{\beta}, \lambda, \kappa)^{\mathrm{T}}$. The solution $\hat{\boldsymbol{\theta}}_{\mathcal{S}}$ satisfy $\sum_{i=1}^{n} \boldsymbol{\phi}(Y_i, D_i, \boldsymbol{X}_{\mathcal{S}i}; \hat{\boldsymbol{\theta}}_{\mathcal{S}}) = 0$, then the first element of $\hat{\boldsymbol{\theta}}_{\mathcal{S}}$ is our IPW estimator (2). We calculate $V_{\mathcal{S}}(\boldsymbol{\theta_0})$ to get the variance of $\hat{\Delta}_{\mathcal{S}}$. We know that the true value $\boldsymbol{\theta}_0 = (\Delta_0, \boldsymbol{\beta}, 1, 1)$. Based on the calculation, we have

$$A_{\mathcal{S}} = E\{-\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\boldsymbol{\phi}(Y, D, \boldsymbol{X}_{\mathcal{S}}; \boldsymbol{\theta}) \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\} = \begin{bmatrix} 1 & H_{\boldsymbol{\beta}}^{\mathrm{T}} & -\mu_1 & \mu_0 \\ 0 & E_{\boldsymbol{\beta}\boldsymbol{\beta}} & 0 & 0 \\ 0 & -E(\frac{1}{e}\frac{\partial e}{\partial \boldsymbol{\beta}^{\mathrm{T}}}) & 1 & 0 \\ 0 & -E(\frac{1}{1-e}\frac{\partial e}{\partial \boldsymbol{\beta}^{\mathrm{T}}}) & 0 & 1 \end{bmatrix},$$

$$B_{\mathcal{S}} = E\{\boldsymbol{\phi}(Y, D, \boldsymbol{X}_{\mathcal{S}}; \boldsymbol{\theta})\boldsymbol{\phi}(Y, D, \boldsymbol{X}_{\mathcal{S}}; \boldsymbol{\theta})^{\mathrm{T}} \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\} =$$

$$\begin{bmatrix} \sigma^2 & H_{\boldsymbol{\beta}}^{\mathrm{T}} & \Delta_0 - E\{\frac{Y(1)}{e}\} & \Delta_0 + E(\frac{Y(0)}{1-e}) \\ H_{\boldsymbol{\beta}} & E_{\boldsymbol{\beta},\boldsymbol{\beta}} & -E(\frac{1}{e}\frac{\partial e}{\partial \boldsymbol{\beta}}) & E(\frac{1}{1-e}\frac{\partial e}{\partial \boldsymbol{\beta}}) \\ \Delta_0 - E\{\frac{Y(1)}{e}\} & -E(\frac{1}{e}\frac{\partial e}{\partial \boldsymbol{\beta}}) & E(\frac{1}{e} - 1) & -1 \\ \Delta_0 + E\{\frac{Y(0)}{1-e}\} & E(\frac{1}{1-e}\frac{\partial e}{\partial \boldsymbol{\beta}}) & -1 & E(\frac{1}{1-e}) - 1 \end{bmatrix},$$

where

$$H = E[\{\frac{Y(1)}{e} + \frac{Y(0)}{1-e}\}\frac{\partial e}{\partial \beta}], \quad \mu_1 = E\{Y(1)\}, \quad \mu_0 = E\{Y(0)\},$$

$$\sigma^2 = E\{\frac{Y(1)^2}{e} + \frac{Y(0)^2}{1-e}\},\ e = e(\boldsymbol{X}_{\mathcal{S}}; \boldsymbol{\beta}) = P(D = 1 \mid \boldsymbol{X}_{\mathcal{S}}),\ E_{\boldsymbol{\beta},\boldsymbol{\beta}} = E\{\frac{1}{e(1-e)}\frac{\partial e}{\partial \boldsymbol{\beta}}\frac{\partial e}{\partial \boldsymbol{\beta}^{\mathrm{T}}}\}.$$

Since we are interested in the asymptotic variance of $\hat{\boldsymbol{\Delta}}_{\mathcal{S}}$, we need the first element of matrix $V_{\mathcal{S}} = A_{\mathcal{S}}^{-1} B_{\mathcal{S}} \{A_{\mathcal{S}}^{-1}\}^{\mathrm{T}}$. The calculation is a little bit complicated thus we omit it. The result be

$$\sigma_0^2 = \sigma_{\mathcal{S}}^2 - H_{\mathcal{S}}^{\mathrm{T}} E_{\boldsymbol{\beta},\boldsymbol{\beta}}^{-1} H_{\mathcal{S}},$$

where

$$H_{\mathcal{S}} = E[\{\frac{Y(1) - \mu_1}{e} + \frac{Y(0) - \mu_0}{1-e}\}\frac{\partial e}{\partial \boldsymbol{\beta}}], \quad \sigma_{\mathcal{S}}^2 = E\{\frac{\{Y(1) - \mu_1\}^2}{e} + \frac{\{Y(0) - \mu_0\}^2}{1-e}\}.$$

To simplify our notation, we denote $H_{\mathcal{S}}$ as $H_{\boldsymbol{\beta}}$.

Similarly, we can write estimation equation $\boldsymbol{\phi}(Y, D, \boldsymbol{X}_{\mathcal{S}_P}; \boldsymbol{\theta}_{\mathcal{S}_P})$ when we want to estimate $\tilde{e}(\boldsymbol{X}_{\mathcal{S}_P}; \boldsymbol{\beta}, \boldsymbol{\gamma})$. We only need to add a group of equations to estimate $\boldsymbol{\gamma}$ which is the coefficient of precision variables $X_{\mathcal{P}'}$. The equations we add into (40) are $\boldsymbol{\phi} = (\frac{D}{\widetilde{e}} - \frac{1-D}{1-\widetilde{e}})\partial\widetilde{e}/\partial\boldsymbol{\gamma}$,

$$\begin{cases} \phi_0 = -\Delta_{\mathcal{S}_P} + \lambda\frac{YD}{\widetilde{e}} - \kappa\frac{Y(1-D)}{1-\widetilde{e}} \\ \phi_i = (\frac{D}{\widetilde{e}} - \frac{1-D}{1-\widetilde{e}})\partial\widetilde{e}/\beta_i \\ \phi_{d+j} = (\frac{D}{\widetilde{e}} - \frac{1-D}{1-\widetilde{\widetilde{e}}})\partial\widetilde{e}/\partial\gamma_j \\ \phi_{p_0+1} = -\lambda\frac{D}{\widetilde{e}} + 1 \\ \phi_{p_0+2} = -\kappa\frac{1-D}{1-\widetilde{e}} + 1, \end{cases} \tag{41}$$

where $1 \le i \le d$, $1 \le j \le p_0 - d$, β_1 be the intercept, β_{i+1} be the coefficient of i'th component of $\boldsymbol{X}_{\mathcal{S}_P}$, γ_j be the coefficient of j'th component of $\boldsymbol{X}_{\mathcal{P}'}$, $\widetilde{e} = \widetilde{e}(\boldsymbol{X}_{\mathcal{S}_P}; \boldsymbol{\beta}, \boldsymbol{\gamma}) = P(D = 1 \mid \boldsymbol{X}_{\mathcal{S}_P})$. We know the coefficient $\boldsymbol{\gamma}$ which correspondent to precision variables is $\mathbf{0}$, but we still estimate it in practice in order to improve efficiency. We write these equations to get the solution of M-estimator $\boldsymbol{\theta}_{\mathcal{S}_P} = (\Delta_{\mathcal{S}_P}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda, \kappa)$, and the true value be $\boldsymbol{\theta}_0 = (\Delta_0, \boldsymbol{\beta}, 0, 1, 1)$. Repeat the process above we can calculate $A_{\mathcal{S}_P}$, $B_{\mathcal{S}_P}$, $V_{\mathcal{S}_P}$. Finally, we find

$$\sigma_P^2 = \sigma_{\mathcal{S}_P}^2 - H_{\mathcal{S}_P}^{\mathrm{T}} E_{\boldsymbol{\beta\gamma},\boldsymbol{\beta\gamma}}^{-1} H_{\mathcal{S}_P},$$

where

$$H_{\mathcal{S}_P} = E[\{\frac{Y(1)-\mu_1}{\widetilde{e}} + \frac{Y(0)-\mu_0}{1-\widetilde{e}}\}\frac{\partial\widetilde{e}}{\partial\boldsymbol{\beta}}, \{\frac{Y(1)-\mu_1}{\widetilde{e}} + \frac{Y(0)-\mu_0}{1-\widetilde{e}}\}\frac{\partial\widetilde{e}}{\partial\gamma}]$$

$$\sigma_{\mathcal{S}_P}^2 = E[\frac{\{Y(1)-\mu_1\}^2}{\widetilde{e}} + \frac{\{Y(0)-\mu_0\}^2}{1-\widetilde{e}}], \quad E_{\boldsymbol{\beta\gamma},\boldsymbol{\beta\gamma}} = \begin{bmatrix} E\{\frac{1}{\widetilde{e}(1-\widetilde{e})}\frac{\partial\widetilde{e}}{\partial\boldsymbol{\beta}}\frac{\partial\widetilde{e}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\} & E\{\frac{1}{\widetilde{e}(1-\widetilde{e})}\frac{\partial\widetilde{e}}{\partial\gamma}\frac{\partial\widetilde{e}}{\partial\boldsymbol{\beta}}\} \\ E\{\frac{1}{\widetilde{e}(1-\widetilde{e})}\frac{\partial\widetilde{e}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\frac{\partial\widetilde{e}}{\partial\gamma}\} & E\{\frac{1}{\widetilde{e}(1-\widetilde{e})}\frac{\partial\widetilde{e}}{\partial\gamma}\frac{\partial\widetilde{e}}{\partial\gamma}\} \end{bmatrix}.$$

We have showed in (32), $D \perp\!\!\!\perp \boldsymbol{X}_{\mathcal{P}'} \mid X_{\mathcal{S}}$, thus

$$e = e(\boldsymbol{X}_{\mathcal{S}}; \boldsymbol{\beta}) = P(D = 1 \mid \boldsymbol{X}_{\mathcal{S}}) = P(D = 1 \mid \boldsymbol{X}_{\mathcal{S}_p}) = \widetilde{e}(\boldsymbol{X}_{\mathcal{S}_P}; \boldsymbol{\beta}, \boldsymbol{\gamma}) = \widetilde{e}. \tag{42}$$

By Matrix partition, we denote $H_{\mathcal{S}_P}$, $E_{\boldsymbol{\beta\gamma},\boldsymbol{\beta\gamma}}$ as

$$(H_{\boldsymbol{\beta}}, H_{\boldsymbol{\gamma}}), \quad \begin{pmatrix} E_{\boldsymbol{\beta},\boldsymbol{\beta}} & E_{\boldsymbol{\gamma},\boldsymbol{\beta}} \\ E_{\boldsymbol{\gamma},\boldsymbol{\beta}}^{\mathrm{T}} & E_{\boldsymbol{\gamma},\boldsymbol{\gamma}} \end{pmatrix},$$

respectively. Again by (42) and matrix calculation, the following result would show immediately,

$$\sigma_0^2 - \sigma_P^2 = (H_{\boldsymbol{\gamma}} - E_{\boldsymbol{\gamma},\boldsymbol{\beta}}^{\mathrm{T}} E_{\boldsymbol{\beta},\boldsymbol{\beta}}^{-1} H_{\boldsymbol{\beta}})^{\mathrm{T}} (E_{\boldsymbol{\gamma},\boldsymbol{\gamma}} - E_{\boldsymbol{\gamma},\boldsymbol{\beta}}^{\mathrm{T}} E_{\boldsymbol{\beta},\boldsymbol{\beta}}^{-1} E_{\boldsymbol{\gamma},\boldsymbol{\beta}})^{-1} (H_{\boldsymbol{\gamma}} - E_{\boldsymbol{\gamma},\boldsymbol{\beta}}^{\mathrm{T}} E_{\boldsymbol{\beta},\boldsymbol{\beta}}^{-1} H_{\boldsymbol{\beta}}) \geq 0.$$

Thus we finished the proof in this part. $\square$

Part B For estimator (1), the procedure is analogous. Thus we only write down the estimation equations of $e(X_{\mathcal{S}}; \boldsymbol{\beta})$ and $e(X_{\mathcal{S}_p}; \boldsymbol{\beta}, \boldsymbol{\gamma})$, respectively.

For $e(X_{\mathcal{S}}; \boldsymbol{\beta})$, $\boldsymbol{\phi}(Y, D, \boldsymbol{X}_{\mathcal{S}}; \boldsymbol{\theta})$ be

$$\begin{cases} \phi_0 = -\Delta_{\mathcal{S}} + \frac{YD}{e} - \frac{Y(1-D)}{1-e} \\ \phi_i = (\frac{D}{e} - \frac{1-D}{1-e}) \partial e / \beta_i, \end{cases} \tag{43}$$

where $1 \leq i \leq d$, β_1 be the intercept, β_{i+1} be the coefficient of i'th component of $\boldsymbol{X}_{\mathcal{S}}$. We write these equations to get the solution of M-estimator $\boldsymbol{\theta}_{\mathcal{S}} = (\Delta_{\mathcal{S}}, \boldsymbol{\beta})^{\mathrm{T}}$. $\hat{\boldsymbol{\theta}}_{\mathcal{S}}$ satisfy

$$\sum_{i=1}^{n} \boldsymbol{\phi}(Y_i, D_i, \boldsymbol{X}_{\mathcal{S}i}; \hat{\boldsymbol{\theta}}_{\mathcal{S}}) = 0.$$

The first element of $\hat{\boldsymbol{\theta}}_{\mathcal{S}}$ is our IPW estimator (1). We know that the true value $\boldsymbol{\theta}_0 = (\Delta_0, \boldsymbol{\beta})$. Based on the calculation

$$\sigma_0^2 = \sigma_{\mathcal{S}}^2 - H_{\mathcal{S}}^{\mathrm{T}} E_{\boldsymbol{\beta},\boldsymbol{\beta}}^{-1} H_{\mathcal{S}},$$

where

$$H_{\mathcal{S}} = E\{(\frac{Y(1)}{e} + \frac{Y(0)}{1-e})\frac{\partial e}{\partial \boldsymbol{\beta}}\}, \quad \sigma^2_{\mathcal{S}} = E[\frac{Y(1)^2}{e} + \frac{Y(0)^2}{1-e}].$$

To simplify notation we use $H_{\boldsymbol{\beta}}$ to denote $H_{\mathcal{S}}$.

Similarly we can write estimate equation $\boldsymbol{\phi}(Y, D, \boldsymbol{X}_{\mathcal{S}_P}; \boldsymbol{\theta}_{\mathcal{S}_P})$ when we estimate $\widetilde{e}(\boldsymbol{X}_{\mathcal{S}_P}; \boldsymbol{\beta}, \boldsymbol{\gamma})$. The equations we add into (43) are $\boldsymbol{\phi} = (\frac{D}{\widetilde{e}} - \frac{1-D}{1-\widetilde{e}})\partial\widetilde{e}/\partial\boldsymbol{\gamma}$. These equations are

$$\begin{cases} \phi_0 = -\Delta_{\mathcal{S}_P} + \frac{YD}{\widetilde{e}} - \frac{Y(1-D)}{1-\widetilde{e}} \\ \phi_i = (\frac{D}{\widetilde{e}} - \frac{1-D}{1-\widetilde{e}})\partial\widetilde{e}/\beta_i \\ \phi_{d+j} = (\frac{D}{\widetilde{e}} - \frac{1-D}{1-\widetilde{e}})\partial\widetilde{e}/\partial\gamma_j, \end{cases}$$

where $1 \le i \le d$, $1 \le j \le p_0 - d$, β_1 correspondent to intercept, β_{i+1} be the coefficient of i'th component of $\boldsymbol{X}_{\mathcal{S}}$. γ_j be the coefficient of j'th component of $\boldsymbol{X}_{\mathcal{P}'}$, $\widetilde{e} = \widetilde{e}(\boldsymbol{X}_{\mathcal{S}_P}; \boldsymbol{\beta}, \boldsymbol{\gamma}) = P(D = 1 \mid \boldsymbol{X}_{\mathcal{S}_P})$. We know that the coefficient $\boldsymbol{\gamma}$ which correspondent to precision variables is $\mathbf{0}$, but we still estimate it in practice in order to improve efficiency. We write these equations to get the solution of M-estimator $\boldsymbol{\theta}_{\mathcal{S}_P} = (\Delta_{\mathcal{S}_P}, \boldsymbol{\beta}, \boldsymbol{\gamma})$. Follow the same argument we used above, we have

$$\sigma^2_P = \sigma^2_{\mathcal{S}_P} - H^{\mathrm{T}}_{\mathcal{S}_P} E^{-1}_{\boldsymbol{\beta\gamma},\boldsymbol{\beta\gamma}} H_{\mathcal{S}_P},$$

where

$$H_{\mathcal{S}_P} = E[\{\frac{Y(1)}{\widetilde{e}} + \frac{Y(0)}{1-\widetilde{e}}\}\frac{\partial\widetilde{e}}{\partial\boldsymbol{\beta}}, \{\frac{Y(1)}{\widetilde{e}} + \frac{Y(0)}{1-\widetilde{e}}\}\frac{\partial\widetilde{e}}{\partial\gamma}], \quad \sigma^2_{\mathcal{M}_{\mathcal{O}}} = E\{\frac{Y(1)^2}{\widetilde{e}} + \frac{Y(0)^2}{1-\widetilde{e}}\} = \sigma^2_{\mathcal{S}},$$

$$E_{\boldsymbol{\beta}\gamma,\boldsymbol{\beta}\gamma} = \begin{bmatrix} E\{\frac{1}{\widetilde{e}(1-\widetilde{e})}\frac{\partial\widetilde{e}}{\partial\boldsymbol{\beta}}\frac{\partial\widetilde{e}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\} & E\{\frac{1}{\widetilde{e}(1-\widetilde{e})}\frac{\partial\widetilde{e}}{\partial\gamma}\frac{\partial\widetilde{e}}{\partial\boldsymbol{\beta}}\} \\ E\{\frac{1}{\widetilde{e}(1-\widetilde{e})}\frac{\partial\widetilde{e}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\frac{\partial\widetilde{e}}{\partial\gamma}\} & E\{\frac{1}{\widetilde{e}(1-\widetilde{e})}\frac{\partial\widetilde{e}}{\partial\gamma}\frac{\partial\widetilde{e}}{\partial\gamma}\} \end{bmatrix}.$$

Similarly, because of (32), we have $D \perp\!\!\!\perp \boldsymbol{X}_{\mathcal{P}'} \mid X_{\mathcal{S}}$, thus

$$e = e(\boldsymbol{X}_{\mathcal{S}}; \boldsymbol{\beta}) = P(D = 1 \mid \boldsymbol{X}_{\mathcal{S}}) = P(D = 1 \mid \boldsymbol{X}_{\mathcal{S}_p}) = \widetilde{e}(\boldsymbol{X}_{\mathcal{S}_P}; \boldsymbol{\beta}, \boldsymbol{\gamma}) = \widetilde{e}.$$

By Matrix partition, we denote $H_{\mathcal{S}_P}$, $E_{\boldsymbol{\beta\gamma},\boldsymbol{\beta\gamma}}$ as

$$(H_{\boldsymbol{\beta}}, H_{\boldsymbol{\gamma}}), \quad \begin{pmatrix} E_{\boldsymbol{\beta},\boldsymbol{\beta}} & E_{\boldsymbol{\gamma},\boldsymbol{\beta}} \\ E^{\mathrm{T}}_{\boldsymbol{\gamma},\boldsymbol{\beta}} & E_{\boldsymbol{\gamma},\boldsymbol{\gamma}} \end{pmatrix},$$

respectively. Again by matrix calculation, the following result would show immediately,

$$\sigma_0^2 - \sigma_P^2 = (H_{\boldsymbol{\gamma}} - E^{\mathrm{T}}_{\boldsymbol{\gamma},\boldsymbol{\beta}} E^{-1}_{\boldsymbol{\beta},\boldsymbol{\beta}} H_{\boldsymbol{\beta}})^{\mathrm{T}} (E_{\boldsymbol{\gamma},\boldsymbol{\gamma}} - E^{\mathrm{T}}_{\boldsymbol{\gamma},\boldsymbol{\beta}} E^{-1}_{\boldsymbol{\beta},\boldsymbol{\beta}} E_{\boldsymbol{\gamma},\boldsymbol{\beta}})^{-1} (H_{\boldsymbol{\gamma}} - E^{\mathrm{T}}_{\boldsymbol{\gamma},\boldsymbol{\beta}} E^{-1}_{\boldsymbol{\beta},\boldsymbol{\beta}} H_{\boldsymbol{\beta}}) \geq 0.$$

Thus we finished the proof of this case. □

8.6 Proof of Theorem 2

Before we prove this result, we need a lemma which showed a subset of instrument variables is independent of the potential outcome given $\boldsymbol{X}_{\mathcal{S}}$:

Lemma If $\boldsymbol{X}_{\mathcal{I}'} \subset \boldsymbol{X}_{\mathcal{I}}$, under assumption 6 we have

$$\boldsymbol{X}_{\mathcal{I}'} \perp\!\!\!\perp Y(d) \mid \boldsymbol{X}_{\mathcal{S}},$$

where d = $0, 1$. Proof for this lemma is straight forward. Combining (36) and contraction property (20), this lemma is an immediate result.

Part A For estimator (2), based on simple calculation and transformation, we have

$$\sqrt{n}(\Delta_{\mathcal{S}_I}-\Delta_0)=\left(n/\sum_{i=1}^{n}\frac{D_i}{\widetilde{e}_i}\right)\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{D_i(Y_i-\mu_1)}{\widetilde{e}_i}\right\}-\left(n/\sum_{i=1}^{n}\frac{1-D_i}{1-\widetilde{e}_i}\right)\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(1-D_i)(Y_i-\mu_0)}{1-\widetilde{e}_i}\right\},$$

$$\sqrt{n}(\Delta_{\mathcal{S}}-\Delta_0)=\left(n/\sum_{i=1}^{n}\frac{D_i}{e_i}\right)\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{D_i(Y_i-\mu_1)}{e_i}\right\}-\left(n/\sum_{i=1}^{n}\frac{1-D_i}{1-e_i}\right)\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(1-D_i)(Y_i-\mu_0)}{1-e_i}\right\}, \tag{44}$$

where

$$e_i=P(D=1|\boldsymbol{X}_{\mathcal{S},i}),\quad \widetilde{e}_i=P(D=1|\boldsymbol{X}_{\mathcal{S}_I,i}),$$

$$\mu_1=E\{Y(1)\},\quad \mu_0=E\{Y(0)\}.$$

We use $q_1^{(n)}$, $q_0^{(n)}$, $\widetilde{q}_1^{(n)}$, $\widetilde{q}_0^{(n)}$ to denote $n/\sum_{i=1}^n(D_i/e_i)$, $\quad n/\sum_{i=1}^n\{(1-D_i)/(1-e_i)\}$, $n/\sum_{i=1}^n(D_i/\widetilde{e}_i)$, $\quad n/\sum_{i=1}^n\{(1-D_i)/(1-\widetilde{e}_i)\}$ respectively. As

$$\begin{aligned}
&E\left(\frac{D}{e}\right)=E\left\{E\left(\frac{D}{e}\mid \boldsymbol{X}_{\mathcal{S}}\right)\right\}=E\left\{\frac{1}{e}E\left(D\mid \boldsymbol{X}_{\mathcal{S}}\right)\right\}=E\left(\frac{e}{e}\right)=1,\\
&E\left(\frac{1-D}{1-e}\right)=E\left\{E\left(\frac{1-D}{1-e}\mid \boldsymbol{X}_{\mathcal{S}}\right)\right\}=E\left\{\frac{1}{1-e}E(1-D|\boldsymbol{X}_{\mathcal{S}})\right\}=E\left\{\frac{1-e}{1-e}\right\}=1,\\
&E\left(\frac{D}{\widetilde{e}}\right)=E\left\{E\left(\frac{D}{\widetilde{e}}\mid \boldsymbol{X}_{\mathcal{S}_I}\right)\right\}=E\left\{\frac{1}{\widetilde{e}}E(D\mid \boldsymbol{X}_{\mathcal{S}_I})\right\}=E\left(\frac{\widetilde{e}}{\widetilde{e}}\right)=1,\\
&E\left(\frac{1-D}{1-\widetilde{e}}\right)=E\left\{E\left(\frac{1-D}{1-\widetilde{e}}\mid \boldsymbol{X}_{\mathcal{S}_I}\right)\right\}=E\left\{\frac{1}{1-\widetilde{e}}E(1-D\mid \boldsymbol{X}_{\mathcal{S}_I})\right\}=E\left\{\frac{1-\widetilde{e}}{1-\widetilde{e}}\right\}=1,
\end{aligned}$$

we have $q_1^{(n)}\xrightarrow{p}1$, $q_0^{(n)}\xrightarrow{p}1$, $\widetilde{q}_1^{(n)}\xrightarrow{p}1$, $\widetilde{q}_0^{(n)}\xrightarrow{p}1$. By central limit theorem, we have the following results:

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{D_i(Y_i-\mu_1)}{e_i} \xrightarrow{d} N\left(0, E\left[\frac{D\{Y(1)-\mu_1\}^2}{e_i^2}\right]\right),\\
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(1-D_i)(Y_i-\mu_0)}{1-e_i} \xrightarrow{d} N\left(0, E\left[\frac{(1-D)\{Y(0)-\mu_0\}^2}{(1-e_i)^2}\right]\right),\\
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{D_i(Y_i-\mu_1)}{e_i}-\frac{(1-D_i)(Y_i-\mu_0)}{1-e_i}\right\} \xrightarrow{d}\\
&N\left(0, E\left[\frac{D\{Y(1)-\mu_1\}^2}{e_i^2}+\frac{(1-D)\{Y(0)-\mu_0\}^2}{(1-e_i)^2}\right]\right).
\end{aligned}
\tag{45}
$$

Meanwhile, we can rewrite $\sqrt{n}(\Delta_{\mathcal{S}}-\Delta_0)$ into the following form,

$$
\begin{aligned}
\sqrt{n}(\Delta_{\mathcal{S}}-\Delta_0) = &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{D_i(Y_i-\mu_1)}{e_i}-\frac{(1-D_i)(Y_i-\mu_0)}{1-e_i}\right\}+\\
&(q_1^{(n)}-1)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{D_i(Y_i-\mu_1)}{e_i}-(q_0^{(n)}-1)\sum_{i=1}^{n}\frac{(1-D_i)(Y_i-\mu_0)}{1-e_i}.
\end{aligned}
\tag{46}
$$

Combining (45), (46) and Slutsky's theorem, we can get the asymptotic distribution of $\sqrt{n}(\Delta_{\mathcal{S}}-\Delta_0)$:

$$
\sqrt{n}(\Delta_{\mathcal{S}}-\Delta_0) \xrightarrow{d} N(0,\sigma_0^2), \tag{47}
$$

where

$$
\begin{aligned}
\sigma_0^2 &= E\left[\frac{D\{Y(1)-\mu_1\}^2}{e^2}+\frac{(1-D)\{Y(0)-\mu_0\}^2}{(1-e)^2}\right]\\
&= E\left[E\left[\frac{\{Y(1)-\mu_1\}^2}{e^2}\mid \boldsymbol{X}_{\mathcal{S}}\right]E(D\mid \boldsymbol{X}_{\mathcal{S}})+E\left[\frac{(Y(0)-\mu_0)^2}{(1-e)^2}\mid \boldsymbol{X}_{\mathcal{S}}\right]E\{(1-D)\mid \boldsymbol{X}_{\mathcal{S}}\}\right]\\
&= E\left[E\left[\frac{\{Y(1)-\mu_1\}^2}{e^2}\mid \boldsymbol{X}_{\mathcal{S}}\right]e+E\left[\frac{\{Y(0)-\mu_0\}^2}{(1-e)^2}\mid \boldsymbol{X}_{\mathcal{S}}\right](1-e)\right]\\
&= E(\frac{(Y(1)-\mu_1)^2}{e}+\frac{(Y(0)-\mu_0)^2}{1-e}).
\end{aligned}
$$

The second equality holds since $X_{\mathcal{S}}$ is a sufficient set. For a similar reason, we have

$$\sqrt{n}(\Delta_{\mathcal{S}_I} - \Delta_0) \xrightarrow{d} N(0, \sigma_I^2),$$

where

$$\sigma_I^2 = E\left[\frac{\{Y(1)-\mu_1\}^2}{\widetilde{e}} + \frac{\{Y(0)-\mu_0\}^2}{1-\widetilde{e}}\right].$$

Now we are ready to show $\sigma_I^2 \geq \sigma_0^2$. Since $\mathcal{I}' \subset \mathcal{I}$, use the lemma we proved at the beginning, we have $E[\{Y(d)-\mu_d\}^2 \mid \boldsymbol{X}_{\mathcal{S}_I}] = E[\{Y(d)-\mu_d\}^2 \mid \boldsymbol{X}_{\mathcal{S}}]$ for $d = 0, 1$. Denote these conditional expectations as $\widetilde{Y}(d)$ while $d = 0, 1$, we have the following results:

$$\begin{aligned}
\sigma_I^2 &= E\left[E\left[\frac{\{Y(1)-\mu_1\}^2}{\widetilde{e}} + \frac{\{Y(0)-\mu_0\}^2}{1-\widetilde{e}} \mid \boldsymbol{X}_{\mathcal{S}_I}\right]\right] \\
&= E\left[\frac{1}{\widetilde{e}}E[\{Y(1)-\mu_1\}^2 \mid \boldsymbol{X}_{\mathcal{S}_I}]\right] + E\left[\frac{1}{1-\widetilde{e}}E[\{Y(0)-\mu_0\}^2 \mid \boldsymbol{X}_{\mathcal{S}_I}]\right] \\
&= E\left\{\frac{1}{\widetilde{e}}\widetilde{Y}(1)\right\} + E\left\{\frac{1}{1-\widetilde{e}}\widetilde{Y}(0)\right\} \\
&= E\left[E\left\{\frac{1}{\widetilde{e}}\widetilde{Y}(1) \mid \boldsymbol{X}_{\mathcal{S}}\right\}\right] + E\left[E\left\{\frac{1}{1-\widetilde{e}}\widetilde{Y}(0) \mid \boldsymbol{X}_{\mathcal{S}}\right\}\right] \\
&= E\left\{\widetilde{Y}(1)E\left(\frac{1}{\widetilde{e}} \mid \boldsymbol{X}_{\mathcal{S}}\right)\right\} + E\left\{\widetilde{Y}(0)E\left(\frac{1}{1-\widetilde{e}} \mid \boldsymbol{X}_{\mathcal{S}}\right)\right\} \\
&\geq E\left\{\widetilde{Y}(1)\frac{1}{e} + \widetilde{Y}(0)\frac{1}{1-e}\right\} \\
&= E\left[\frac{1}{e}E[\{Y(1)-\mu_1\}^2 \mid \boldsymbol{X}_{\mathcal{S}}] + \frac{1}{1-e}E[\{Y(0)-\mu_0\}^2 \mid \boldsymbol{X}_{\mathcal{S}}]\right] \\
&= E\left[E\left[\frac{\{Y(1)-\mu_1\}^2}{e} + \frac{\{Y(0)-\mu_0\}^2}{1-e} \mid \boldsymbol{X}_{\mathcal{S}}\right]\right] \\
&= \sigma_0^2.
\end{aligned}$$

The inequality holds since

$$1 = E(1 \mid \boldsymbol{X}_{\mathcal{S}}) = E\left(\frac{1}{\widetilde{e}}\widetilde{e} \mid \boldsymbol{X}_{\mathcal{S}}\right) \leq E\left(\frac{1}{\widetilde{e}} \mid \boldsymbol{X}_{\mathcal{S}}\right) E\left(\widetilde{e} \mid \boldsymbol{X}_{\mathcal{S}}\right) = E\left(\frac{1}{\widetilde{e}} \mid \boldsymbol{X}_{\mathcal{S}}\right) e.$$

For the same reason,

$$(1-e)E(\frac{1}{1-\widetilde{e}}|\boldsymbol{X}_{\mathcal{S}}) \geq 1.$$

Thus we finished our proof for this case.

Part B Based on transformation, Horvitz-Thompson estimator (1) could be rewritten into following form:

$$\sqrt{n}(\Delta_{\mathcal{S}} - \Delta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{D_i Y_i}{\widetilde{e}_i} - \frac{(1-D_i)Y_i}{1-\widetilde{e}_i} - \Delta_0 \right\},$$

$$\sqrt{n}(\Delta_{\mathcal{S}_I} - \Delta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{D_i Y_i}{e_i} - \frac{(1-D_i)Y_i}{1-e_i} - \Delta_0 \right\},$$

where $e_i = P(D=1|\boldsymbol{X}_{\mathcal{S}i})$, $\widetilde{e}_i = P(D=1|\boldsymbol{X}_{\mathcal{S}_I i})$. We note that

$$\frac{D_i Y_i}{\widetilde{e}_i} - \frac{(1-D_i)Y_i}{1-\widetilde{e}_i} - \Delta_0, \quad \frac{D_i Y_i}{e_i} - \frac{(1-D_i)Y_i}{1-e_i} - \Delta_0$$

are i.i.d samples, respectively. So from Central limit theorem,

$$\sqrt{n}(\Delta_{\mathcal{S}} - \Delta_0) \xrightarrow{d} N(0, \sigma_0^2),$$

$$\sqrt{n}(\Delta_{\mathcal{S}_I} - \Delta_0) \xrightarrow{d} N(0, \sigma_I^2),$$

where

$$
\begin{aligned}
\sigma_0^2 &= E\left\{\frac{DY(1)^2}{e^2} + \frac{(1-D)Y(0)^2}{(1-e)^2}\right\} - \Delta_0^2 \\
&= E\left[E\left\{\frac{DY(1)^2}{e^2} \mid \boldsymbol{X}_{\mathcal{S}}\right\}\right] + E\left[E\left\{\frac{(1-D)Y(0)^2}{(1-e)^2} \mid \boldsymbol{X}_{\mathcal{S}}\right\}\right] - \Delta_0^2 \\
&= E\left[\frac{1}{e^2}E\{D \mid \boldsymbol{X}_{\mathcal{S}}\}E\{Y(1)^2)|\boldsymbol{X}_{\mathcal{S}}\}\right] + E\left[\frac{1}{(1-e)^2}E\{(1-D) \mid \boldsymbol{X}_{\mathcal{S}}\}E\{Y(0)|\boldsymbol{X}_{\mathcal{S}}\}\right] - \Delta_0^2 \\
&= E\left[\frac{1}{e}E\{Y(1)^2 \mid \boldsymbol{X}_{\mathcal{S}}\}\right] + E\left[\frac{1}{1-e}E\{Y(0)^2|\boldsymbol{X}_{\mathcal{S}}\}\right] - \Delta_0^2 \\
&= E\left[E\left\{\frac{Y(1)^2}{e} + \frac{Y(0)^2}{1-e} \mid \boldsymbol{X}_{\mathcal{S}}\right\}\right] - \Delta_0^2 \\
&= E\left\{\frac{Y(1)^2}{e} + \frac{Y(0)^2}{1-e}\right\} - \Delta_0^2.
\end{aligned}
$$

The third equality holds because $X_{\mathcal{S}}$ is a sufficient set, so $D \perp\!\!\!\perp Y(d) \mid X_{\mathcal{S}}$. For the same reason,

$$
\sigma_I^2 = E\left\{\frac{Y(1)^2}{\widetilde{e}} + \frac{Y(0)^2}{1-\widetilde{e}}\right\} - \Delta_0^2.
$$

Next, we prove $\sigma_I^2 \geq \sigma_0^2$. Since $\mathcal{I}' \subset \mathcal{I}$, together with the lemma we proved at the beginning, we have: $E[Y(d)^2 \mid \boldsymbol{X}_{\mathcal{S}}] = E[Y(d)^2 \mid \boldsymbol{X}_{\mathcal{S}_I}]$ for $d = 0, 1$. Denote these conditional expectation as $\widetilde{Y}(d)$ for $d = 0, 1$ respectively. We have the following results:

$$
\begin{aligned}
\sigma_I^2 &= E\left[E\left\{\frac{Y(1)^2}{\widetilde{e}} + \frac{Y(0)^2}{1-\widetilde{e}} \mid \boldsymbol{X}_{\mathcal{S}_I}\right\}\right] - \Delta_0^2 \\
&= E\left[\frac{1}{\widetilde{e}} E\{Y(1)^2 \mid \boldsymbol{X}_{\mathcal{S}_I}\}\right] + E\left[\frac{1}{1-\widetilde{e}} E\{Y(0)^2 \mid \boldsymbol{X}_{\mathcal{S}_I}\}\right] - \Delta_0^2 \\
&= E\left\{\frac{1}{\widetilde{e}} \widetilde{Y}(1)\right\} + E\left\{\frac{1}{1-\widetilde{e}} \widetilde{Y}(0)\right\} \\
&= E\left[E\left\{\frac{1}{\widetilde{e}} \widetilde{Y}(1) \mid \boldsymbol{X}_{\mathcal{S}}\right\}\right] + E\left[E\left\{\frac{1}{1-\widetilde{e}} \widetilde{Y}(0) \mid \boldsymbol{X}_{\mathcal{S}}\right\}\right] \\
&= E\left\{\widetilde{Y}(1) E\left(\frac{1}{\widetilde{e}} \mid \boldsymbol{X}_{\mathcal{S}}\right)\right\} + E\left\{\widetilde{Y}(0) E\left(\frac{1}{1-\widetilde{e}} \mid \boldsymbol{X}_{\mathcal{S}}\right)\right\} \\
&\geq E\left\{\widetilde{Y}(1) \frac{1}{e} + \widetilde{Y}(0) \frac{1}{1-e}\right\} \\
&= E\left[\frac{1}{e} E\{Y(1)^2 \mid \boldsymbol{X}_{\mathcal{S}}\} + \frac{1}{1-e} E\{Y(0)^2 \mid \boldsymbol{X}_{\mathcal{S}}\}\right] \\
&= E\left[E\left\{\frac{Y(1)^2}{e} + \frac{Y(0)^2}{1-e} \mid \boldsymbol{X}_{\mathcal{S}}\right\}\right] = \sigma_0^2.
\end{aligned}
$$

The reason why inequality holds is showed in proof of part (A). So far, we have already accomplished the proof of theorem 2. $\square$.

8.7 Proof of Theorem 3

We firstly show that there exists a constant $\tilde{c}$ such that

$$
P(|\alpha - \hat{\alpha}| > cn^{-k}) \leq O(\exp(-\tilde{c}n^{1-2k})), \tag{48}
$$

where $c > 0$ is a constant, $0 < \kappa < 1/2$.

Define $\alpha = BCov^2(X, Y|D)$, $\hat{\alpha} = BCov_n^2(X, Y|D)$, $\alpha_1 = BCov^2(X^{(1)}, Y^{(1)})$ and $\alpha_0 = BCov^2(X^{(0)}, Y^{(0)})$, where $X^{(d)}$, $Y^{(d)}$ follow the same distributions as $(X \mid D = d)$, $(Y \mid D = d)$ for $d = 0, 1$ respectively. We use $\hat{\alpha}_1$ $\hat{\alpha}_0$ to denote their sample estimators $BCov_{n_1}(X^{(1)}, Y^{(1)})$, $BCov_{n_0}(X^{(0)}, Y^{(0)})$, respectively. We divide the sample $(Y_i, D_i, \boldsymbol{X})_{i=1}^n$ into two parts $(\boldsymbol{X}_i^{(1)}, Y_i^{(1)})_{i=1}^{n_1}$ and $(\boldsymbol{X}_i^{(0)}, Y_i^{(1)})_{i=1}^{n_0}$, and the in the first part the value of D_i always be 1, while the other part the value of D_i always be 0. Recall $n_1 = \sum_{i=1}^n D_i$, $n_0 = n - n_1$, $\omega = P(D=1)$ and $\hat{\omega} = n_1/n$.

We can write

$$\begin{aligned}\alpha - \hat{\alpha} &= \omega\alpha_1 + (1-\omega)\alpha_0 - (\hat{\omega}\hat{\alpha}_1 + (1-\hat{\omega})\hat{\alpha}_0) \\ &= \omega(\alpha_1 - \hat{\alpha}_1) + (1-\omega)(\alpha_0 - \hat{\alpha}_0) + (\hat{\alpha}_1 - \hat{\alpha}_0)(\omega - \hat{\omega}).\end{aligned} \tag{49}$$

Since $\alpha_1, \alpha_0, \hat{\alpha}_1, \hat{\alpha}_0 \in [0,1]$, $|\alpha - \hat{\alpha}| \le \omega|\alpha_1 - \hat{\alpha}_1| + (1-\omega)|\alpha_0 - \hat{\alpha}_0| + |\omega - \hat{\omega}|$. We have

$$\begin{aligned}P(|\alpha - \hat{\alpha}| \ge 2\epsilon) &\le P(\omega|\alpha_1 - \hat{\alpha}_1| \ge \omega\epsilon) + P((1-\omega)|\alpha_0 - \hat{\alpha}_0| \ge (1-\omega)\epsilon) + P(|\omega - \hat{\omega}| \ge \epsilon) \\ &= P(|\alpha_1 - \hat{\alpha}_1| \ge \epsilon) + P(|\alpha_0 - \hat{\alpha}_0| \ge \epsilon) + P(|\omega - \hat{\omega}| \ge \epsilon).\end{aligned} \tag{50}$$

We will deal with the three terms respectively. To begin with, we handle the third term of (50). We note that

$$\omega - \hat{\omega} = \frac{1}{n}\sum_{i=1}^n(\omega - D_i) = \sum_{i=1}^n Z_i,$$

where $Z_i = (\omega - D_i)/n$ be independent zero-mean random variables, and $|Z_i| \le 1/n = M$, $E(Z_i^2) = \omega(1-\omega)/n^2$. Based on the Bernstein inequality, we have

$$P(\omega - \hat{\omega} \ge \epsilon) = P(\sum_{i=1}^n Z_i \ge \epsilon) \le \exp\left(-\frac{\frac{1}{2}\epsilon^2}{\frac{\omega(1-\omega)}{n} + \frac{\epsilon}{3n}}\right)$$

So,

$$P(|\hat{\omega} - \omega| \ge \epsilon) = P(\sum_{i=1}^n Z_i \ge \epsilon) + P(-\sum_{i=1}^n Z_i \ge \epsilon) \le 2\exp\left(-\frac{\frac{1}{2}\epsilon^2}{\frac{\omega(1-\omega)}{n} + \frac{\epsilon}{3n}}\right) \tag{51}$$

Now we turn to the first and second term of (50). Following equation (A.7) from the appendix of Pan et al. (2019a), there exist two positive constants c_1 and c_2 such that

$$\begin{aligned} &P(|\alpha_1 - \hat{\alpha}_1| \geq \epsilon) \leq 2\exp(-c_1 n_1 \epsilon^2), \\ &P(|\alpha_0 - \hat{\alpha}_0| \geq \epsilon) \leq 2\exp(-c_0 n_0 \epsilon^2). \end{aligned} \tag{52}$$

We now show that $\exp(-c_1 n_1 \epsilon^2) = O_P(\exp(-c_1 n\omega\epsilon^2/2))$. We observed that $\omega = P(D = 1) > 0$, we have

$$\begin{aligned} &P\left(\left|\frac{\exp(-c_1 n_1 \epsilon^2)}{\exp(-c_1 \omega n \epsilon^2/2)}\right| > 1\right) \\ &= P(\frac{n\omega}{2} - n_1 > 0) \\ &= P(\omega - \hat{\omega} > \frac{\omega}{2}) \\ &= P(\sum_{i=1}^{n} Z_i > \frac{\omega}{2}) \\ &\leq \exp\left(-\frac{\frac{1}{8}\omega^2}{\frac{\omega(1-\omega)}{n} + \frac{\omega}{6n}}\right) \\ &\overset{n\to\infty}{\longrightarrow} 0. \end{aligned} \tag{53}$$

For the same reason, we have

$$\exp(-c_0 n_0 \epsilon^2) = O_P(\exp(-c_0 n(1-\omega)\epsilon^2/2)). \tag{54}$$

If $\epsilon < 3\omega(1-\omega)$, we have

$$\exp\left(-\frac{\frac{1}{2}\epsilon^2}{\frac{\omega(1-\omega)}{n} + \frac{\epsilon}{3n}}\right) = \exp\left(-\frac{1}{2\omega(1-\omega) + 2\epsilon/3} n\epsilon^2\right) \leq \exp(-\tilde{c}_2 n \epsilon^2), \tag{55}$$

where $\tilde{c}_2 = 1/\{4\omega(1-\omega)\}$. Combining (51), (53), (54), (55). Let $\epsilon = cn^{-\kappa}/2$, where $0 < \kappa <$

$1/2$, $\tilde{c}_1 = c_1\omega/2$, $\tilde{c}_0 = c_0(1-\omega)/2$, we have

$$P(|\alpha - \hat{\alpha}| \geq cn^{-\kappa}) \leq O_P(\exp(-\tilde{c}_1 cn^{1-2\kappa})) + O_P(\exp(-\tilde{c}_0 cn^{1-2\kappa}))$$
$$+ O_P(\exp(-\tilde{c}_2 cn^{1-2\kappa}))$$

Let $\tilde{c} = min(c\tilde{c}_1, c\tilde{c}_0, c\tilde{c}_2)$, then we have

$$P(|\alpha - \hat{\alpha}| \geq cn^{-\kappa}) \leq O_P(\exp(-\tilde{c}n^{1-2\kappa}))$$

Hence we finished the proof of equation (48). Now let $\rho_j = BCov^2(X_i, Y|D)$ and $\hat{\rho}_j = BCov_n^2(X_i, Y|D)$ for j = $1, 2, \ldots, p$, from equation (48) we know that $P(|\hat{\rho}_j - \rho_j| > cn^{-\kappa}) \leq O(exp(-c_1 n^{1-2\kappa}))$.

As $\tau_n \in (0, cn^{-\kappa})$ and $\{(X_{\mathcal{C}} \cup X_{\mathcal{P}}) \not\subset \hat{A}_n^*\} \subset \{|\hat{\rho}_j - \rho_j| > cn^{-\kappa}, \text{ for some j } \in (X_{\mathcal{C}} \cup X_{\mathcal{P}})\}$, we have

$$P(\{(X_{\mathcal{C}} \cup X_{\mathcal{P}}) \subset \hat{A}_n^*\}) \geq 1 - \eta P(|\hat{\rho}_j - \rho_j| > cn^{-\kappa}) \geq 1 - \eta O(\exp(-\tilde{c}n^{1-2\kappa})),$$

where η is the cardinality of $(X_{\mathcal{C}} \cup X_{\mathcal{P}})$. Hence

$$P(\{(X_{\mathcal{C}} \cup X_{\mathcal{P}}) \subset \hat{A}_n^*\}) \stackrel{n\to\infty}{\longrightarrow} 1.$$

8.8 Proof of Theorem 4

Part c Before we start, we need a technical lemma, which tells us under some conditions, an M-estimator is a consistent estimator. The proof of the lemma could be found at van der Vaart (1998), page 46, theorem 5.9.

Lemma 1. *Let Φ_n be random vector-valued functions and let Φ be a fixed vector function of $\boldsymbol{\theta}$ such that for every $\epsilon > 0$, following conditions hold:*

(i)$\sup_{\boldsymbol{\theta}\in\Theta} \|\Phi_n(\boldsymbol{\theta}) - \Phi(\boldsymbol{\theta})\| \xrightarrow{p} \mathbf{0}$,

(ii)$\inf_{\boldsymbol{\theta}:d(\boldsymbol{\theta},\theta_0)\geq\epsilon} \|\Phi(\boldsymbol{\theta})\| > 0 = \|\Phi(\boldsymbol{\theta}_0)\|$

Then any sequence of estimators $\hat{\boldsymbol{\theta}}_n$ such that $\Phi_n(\hat{\boldsymbol{\theta}}_n) = o_p(1)$ converge in probability to $\boldsymbol{\theta}_0$.

Lemma 2.

For the proof, we assume the following regularity conditions:

1. We assume all expectations exist and finite.

2. We assume all estimation equations Φ and Φ_n satisfy condition (i) (ii) of the lemma. This assumption is mild since we always expect an M-estimator is a consistent estimator.

We use $\boldsymbol{\beta}^*$ to denote the coefficient estimated by our procedure in Section 4.2, $\hat{\boldsymbol{\beta}}$ to denote the coefficient estimation when we use $\mathcal{A}$ as prior. Let $e_i^* = e(\boldsymbol{X}_i; \boldsymbol{\beta}^*)$, $\hat{e}_i = e(\boldsymbol{X}_i; \hat{\boldsymbol{\beta}})$. We use Δ_{HT}, Δ_{Ratio}, Δ_{DR} to denote IPW estimators: (1), (2), (3) respectively. Without loss of generality we assume $\mathcal{A} = \{1, 2, 3, ..., p_0\}$. We aim to prove the following results:

$$\sqrt{n}(\Delta^*_{HT} - \hat{\Delta}_{HT}) \xrightarrow{p} 0,$$
$$\sqrt{n}(\Delta^*_{Ratio} - \hat{\Delta}_{Ratio}) \xrightarrow{p} 0,$$
$$\sqrt{n}(\Delta^*_{DR} - \hat{\Delta}_{DR}) \xrightarrow{p} 0,$$

where we use plug-in estimator e_i^*, $\hat{e}_i$ to construct IPW estimator using (1), (2), (3), respectively.

To begin with, we are going to show

$$\sqrt{n}(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) \xrightarrow{p} 0. \tag{56}$$

From part (a) we know that for any $j \notin \mathcal{A}$, $\lim_{n\to\infty} P(\beta_j^* \neq 0) = 0$. And when we estimate $\boldsymbol{\beta}$, we only use variables in $\mathcal{A}$, thus $\hat{\beta}_j \equiv 0$. We have shown that

$$\lim_{n\to\infty} P(\boldsymbol{\beta}^*_{\mathcal{A}^c} - \hat{\boldsymbol{\beta}}_{\mathcal{A}^c} \neq 0) = 0,$$

which implies $\sqrt{n}(\boldsymbol{\beta}^*_{\mathcal{A}^c} - \hat{\boldsymbol{\beta}}_{\mathcal{A}^c}) \xrightarrow{p} 0$. Now we are going to show $\sqrt{n}(\boldsymbol{\beta}^*_{\mathcal{A}} - \hat{\boldsymbol{\beta}}_{\mathcal{A}}) \xrightarrow{d} 0$. By the KKT conditions, we must have

$$
\begin{aligned}
&|\sum_{i=i}^{n} X_{ij}\left\{D_i - e(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}^*)\right\}| \le \frac{\lambda_n}{\hat{\omega}_j^{(n)}}, \\
&\sum_{i=i}^{n} X_{ij}\left\{D_i - e(\boldsymbol{X}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}})\right\} = 0,
\end{aligned}
\tag{57}
$$

where $j \in \mathcal{A}$, $e = e(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})$ is the $Logistic$ model we specified before. We make a subtraction:

$$
\frac{1}{\sqrt{n}}|\sum_{i=1}^{n} X_{ij}\{e(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}^*) - e(\boldsymbol{X}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}})\}| \le \frac{\lambda_n}{\sqrt{n}\hat{\omega}_j^{(n)}}. \tag{58}
$$

We use $\boldsymbol{\beta}_0$ to denote the coefficient of the oracle propensity score model, which is $e(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0) = P(D = 1 \mid \boldsymbol{X}_{\mathcal{A}})$. Let

$$
\begin{aligned}
&\boldsymbol{\beta}^* = \boldsymbol{\beta}_0 + \frac{\boldsymbol{u}^*}{\sqrt{n}}, \\
&\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \frac{\hat{\boldsymbol{u}}}{\sqrt{n}}, \\
&\boldsymbol{u} = \boldsymbol{u}^* - \hat{\boldsymbol{u}}.
\end{aligned}
$$

We are going to make a Taylor expansion for (58) at point $\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0$

$$
\begin{aligned}
e(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}^*) &= e(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0) + e'(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0)\frac{\boldsymbol{u}^*}{\sqrt{n}} + e''(U_i)\frac{(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{u}^*)^2}{n}, \\
e(\boldsymbol{X}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}) &= e(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0) + e'(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0)\frac{\hat{\boldsymbol{u}}}{\sqrt{n}} + e''(V_i)\frac{(\boldsymbol{X}_i^{\mathrm{T}}\hat{\boldsymbol{u}})^2}{n},
\end{aligned}
$$

where U_i is between $\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}^*$ and $\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0$, V_i is between $\boldsymbol{X}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}$ and $\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0$. The left side of (58) can be written as $A_1^{(n)} + A_2^{(n)}$, where

$$A_1^{(n)} = \sum_{i=1}^{n} \frac{X_{ij}}{n} e'(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}_0) \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u},$$
$$A_2^{(n)} = \sum_{i=1}^{n} \frac{X_{ij}}{n^{3/2}} \left\{ e''(U_i)(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u}^*)^2 - e''(V_i)(\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{u}})^2 \right\},$$

We can rewrite (58) into vector form:

$$\frac{\lambda_n}{\sqrt{n}} \boldsymbol{w} \geq |\boldsymbol{A}_{\boldsymbol{1}}^{(n)} + \boldsymbol{A}_{\boldsymbol{2}}^{(n)}|, \tag{59}$$
$$\boldsymbol{A}_1^{(n)} = \frac{1}{n} \sum_{i=1}^{n} e'(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}_0) \boldsymbol{X}_{i\mathcal{A}} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u},$$
$$\boldsymbol{A}_2^{(n)} = \frac{1}{n^{3/2}} \sum_{i=1}^{n} \boldsymbol{X}_{i\mathcal{A}} \left\{ e''(U_i)(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u}^*)^2 - e''(V_i)(\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{u}})^2 \right\},$$

where $\boldsymbol{w} = (1/\hat{\omega}_1^{(n)}, \ldots, 1/\hat{\omega}_{p_0}^{(n)})$, $X_{i\mathcal{A}} = (X_{i1}, X_{i2}, \ldots, X_{ip_0})^{\mathrm{T}}$. Let $n \to \infty$. As $\forall j \in \mathcal{A}$, $\hat{\omega}_j^{(n)} \xrightarrow{p} c_j > 0$ and $\lambda_n/\sqrt{n} \xrightarrow{p} 0$, by the Continuous mapping theorem, we must have $\lambda_n \boldsymbol{w}/\sqrt{n} \xrightarrow{p} 0$. We also have

$$\frac{1}{n} \sum_{i=1}^{n} e'(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}_0) \boldsymbol{X}_{i\mathcal{A}} \boldsymbol{X}_i^{\mathrm{T}} \xrightarrow{p} E(e'(\boldsymbol{X}^{\mathrm{T}} \boldsymbol{\beta}_0) \boldsymbol{X}_{\mathcal{A}} \boldsymbol{X}^{\mathrm{T}}).$$

As $\boldsymbol{u}_{\mathcal{A}^c} = \sqrt{n}(\boldsymbol{\beta}^*_{\mathcal{A}^c} - \hat{\boldsymbol{\beta}}_{\mathcal{A}^c}) \xrightarrow{p} \mathbf{0}$, if we can show $\boldsymbol{A}_2^{(n)} \xrightarrow{p} \mathbf{0}$, by Slutsky's theorem, we must have $\boldsymbol{u}_{\mathcal{A}} \xrightarrow{p} \mathbf{0}$. Now we are going to show $\boldsymbol{A}_2^{(n)} \xrightarrow{p} \mathbf{0}$. More precisely, we are going to show that for all $j \in \mathcal{A}$, we have $\sum_{i=1}^{n} X_{ij} e''(V_i)(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u}^*)^2/n^{3/2} \xrightarrow{p} 0$ and $\sum_{i=1}^{n} X_{ij} e''(V_i)(\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{u}})^2/n^{3/2} \xrightarrow{p} 0$.

We note that for *Logistic* model, we have $0 < |e| < 1$, $|e'| = |e(1-e)| < 1$, $|e''| = |e(1-e)(1-2e)| < 1$. We note maximal likely hood estimator is asymptotically normal, so we have $\hat{\boldsymbol{u}} = \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \hat{\Sigma})$, and from part (b) we have $\boldsymbol{u}^* = \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \Sigma^*)$.

We now show that $\sum_{i=1}^{n} X_{ij} e''(U_i)(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u}^*)^2 / n^{3/2} \xrightarrow{p} 0$. We observe that

$$
\begin{aligned}
&\frac{1}{n^{3/2}} \sum_{i=1}^{n} |X_{ij} e''(V_i)(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u}^*)^2| \\
&\leq \sum_{i=1}^{n} \frac{|X_{ij}||\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{X}_i|}{n} \frac{|\boldsymbol{u}^{*\mathrm{T}} \boldsymbol{u}^*|}{n^{1/2}} \\
&\leq \sum_{i=1}^{n} \frac{|X_{ij}||\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{X}_i|}{n} \frac{|\boldsymbol{u}^{*\mathrm{T}}|}{n^{1/4}} \frac{|\boldsymbol{u}^*|}{n^{1/4}}.
\end{aligned}
$$

By the Weak Law of Large Numbers, we have $\sum_{i=1}^{n} |X_{ij}||\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{X}_i|/n \xrightarrow{p} E(|X_i||\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}|) < \infty$. Besides, we have $|\boldsymbol{u}^*|/n^{1/4} \xrightarrow{p} \mathbf{0}$. Thus, by the Continuous mapping theorem, we see that $\sum_{i=1}^{n} X_{ij} e''(U_i)(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u}^*)^2 / n^{3/2} \xrightarrow{p} 0$. Similarly, we have $\sum_{i=1}^{n} X_{ij} e''(V_i)(\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{u}})^2 / n^{3/2} \xrightarrow{p} 0$. So far, we have finished the proof for (56).

Now we are going to show that CBS propensity score estimator, e_i^*, and oracle propensity score estimator, $\hat{e}_i$ are the asymptotic equivalent:

$$
\sqrt{n}(e(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}^*) - e(\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}})\} \xrightarrow{p} 0. \tag{60}
$$

Again, Using the Taylor expansion at point $\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}$, we have $e(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}^*) = e(\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}) +$ $e'(\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}) \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u} / \sqrt{n} + e''(T_i)(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u})^2 / n$, where T_i is between $\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}^*$ and $\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}$. We have

$$
\begin{aligned}
\sqrt{n}|e(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}^*| - e(\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}})| &= \left| e'(\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}) \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u} + \frac{e''(T_i)(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u})^2}{\sqrt{n}} \right| \\
&\leq |e'(\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}) \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u}| + \left| \frac{e''(T_i)(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u})^2}{\sqrt{n}} \right| \\
&\leq |\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u}| + \left| \frac{(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u})^2}{\sqrt{n}} \right| \\
&\leq |\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u}| + \frac{(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{X}_i)(\boldsymbol{u}^{\mathrm{T}} \boldsymbol{u})}{\sqrt{n}}.
\end{aligned}
$$

We use Cauchy inequality to get the last inequality. As $\boldsymbol{u} \xrightarrow{p} 0$, we have $\sqrt{n}|e(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}^*| - e(\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}})| \xrightarrow{p} 0$.

(i) Proof For Horvitz-Thompson Estimator Δ_{HT}.

We have

$$
\begin{aligned}
&\sqrt{n}|\Delta^*_{HT} - \hat{\Delta}_{HT}| \\
&= \left| \frac{\sqrt{n}}{n} \sum_{i=1}^{n} \left\{ Y_i D_i \left(\frac{1}{e_i^*} - \frac{1}{\hat{e}_i} \right) - Y_i(1-D_i) \left(\frac{1}{1-e_i^*} - \frac{1}{1-\hat{e}_i} \right) \right\} \right| \\
&\leq \frac{\sqrt{n}}{n} \sum_{i=1}^{n} \left| \frac{Y_i D_i}{e_i^* \hat{e}_i} - \frac{Y_i(1-D_i)}{(1-e_i^*)(1-\hat{e}_i)} \right| \cdot |\hat{e}_i - e_i^*| \\
&\leq \frac{1}{n} \sum_{i=1}^{n} \left\{ \left| \frac{Y_i D_i}{e_i^{*2}} \right| + \left| \frac{Y_i D_i}{\hat{e}_i^2} \right| + \left| \frac{Y_i(1-D_i)}{(1-e_i^*)^2} \right| + \left| \frac{Y_i(1-D_i)}{(1-\hat{e}_i)^2} \right| \right\} \cdot \sqrt{n}|\hat{e}_i - e_i^*| \\
&\leq \frac{1}{n} \sum_{i=1}^{n} \left\{ \left| \frac{Y_i D_i}{e_i^{*2}} \right| + \left| \frac{Y_i D_i}{\hat{e}_i^2} \right| + \left| \frac{Y_i(1-D_i)}{(1-e_i^*)^2} \right| + \left| \frac{Y_i(1-D_i)}{(1-\hat{e}_i)^2} \right| \right\} \cdot \left(|\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{u}| + \frac{(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{X}_i)(\boldsymbol{u}^{\mathrm{T}} \boldsymbol{u})}{\sqrt{n}} \right),
\end{aligned} \tag{61}
$$

where $e_i^* = e(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}^*)$, $\hat{e}_i = e(\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}})$.

We now show that

$$
\frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i D_i \boldsymbol{X}_{\boldsymbol{i}}^{\mathrm{T}}}{\hat{e}_i^2} \right| \xrightarrow{p} E \left\{ \frac{Y D \boldsymbol{X}^{\mathrm{T}}}{e(\boldsymbol{X} \boldsymbol{\beta}_0)^2} \right\}, \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i D_i \boldsymbol{X}_{\boldsymbol{i}}^{\mathrm{T}}}{e_i^{*2}} \right| \xrightarrow{p} E \left\{ \frac{Y D \boldsymbol{X}^{\mathrm{T}}}{e(\boldsymbol{X}^{\mathrm{T}} \boldsymbol{\beta}_0)^2} \right\}. \tag{62}
$$

We define $L(\boldsymbol{\beta}) = (1/n) \sum_{i=1}^{n} |Y_i D_i \boldsymbol{X}_i^{\mathrm{T}}| / e(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta})^2$, $D(\boldsymbol{\beta}) = L(\boldsymbol{\beta}) - L(\boldsymbol{\beta}_0)$. Assume $\boldsymbol{\beta}$ is a consistent estimator of $\boldsymbol{\beta}_0$, we have

$$
D(\boldsymbol{\beta}) = \left\{ \frac{\partial L(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}_0} + o_p(1) \right\} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^{\mathrm{T}}
$$

So if $\boldsymbol{\beta} \xrightarrow{p} \boldsymbol{\beta}_0$, we have $D(\boldsymbol{\beta}) \xrightarrow{p} 0$. So we have $D(\boldsymbol{\beta}^*) \xrightarrow{p} 0$ and $D(\hat{\boldsymbol{\beta}}) \xrightarrow{p} 0$, which imply $(1/n) \sum_{i=1}^{n} |Y_i D_i \boldsymbol{X}_i^{\mathrm{T}}| / e(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}^*)^2 = L(\boldsymbol{\beta}^*) = L(\boldsymbol{\beta}_0) + D(\boldsymbol{\beta}^*) \xrightarrow{p} E\{|Y D \boldsymbol{X}^{\mathrm{T}}| / e(\boldsymbol{X}^{\mathrm{T}} \boldsymbol{\beta}_0)^2\}$ and $(1/n) \sum_{i=1}^{n} |Y_i D_i \boldsymbol{X}_i^{\mathrm{T}}| / e(\boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}})^2 = L(\hat{\boldsymbol{\beta}}) = L(\boldsymbol{\beta}_0) + D(\hat{\boldsymbol{\beta}}) \xrightarrow{p} E\{|Y D \boldsymbol{X}^{\mathrm{T}}| / e(\boldsymbol{X}^{\mathrm{T}} \boldsymbol{\beta}_0)^2\}$.

For the same reason, the following relationships could be shown analogously:

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i(1-D_i)\boldsymbol{X}_i^{\mathrm{T}}|}{(1-e_i^*)^2}, \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i(1-D_i)\boldsymbol{X}_i^{\mathrm{T}}|}{(1-\hat{e}_i)^2} \xrightarrow{p} E\left[\frac{|Y(1-D)\boldsymbol{X}^{\mathrm{T}}|}{\{1-e(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^2}\right],\\
&\frac{1}{n}\sum_{i=1}^{n}\frac{|Y_iD_i\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{X}_i|}{\hat{e}_i^2}, \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_iD_i\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{X}_i|}{\hat{e}_i^2} \xrightarrow{p} E\left\{\frac{|YD\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}|}{e(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0)^2}\right\},\\
&\frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i(1-D_i)\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{X}|}{(1-e_i^*)^2}, \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i(1-D_i)\boldsymbol{X}\boldsymbol{X}_i^{\mathrm{T}}|}{(1-\hat{e}_i)^2} \xrightarrow{p} E\left[\frac{|Y(1-D)\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}|}{\{1-e(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^2}\right],
\end{aligned}
$$

Given these relationships and $\boldsymbol{u} \xrightarrow{p} 0$, by the Continuous mapping theorem, we can conclude that the right side of (61) $\xrightarrow{p} 0$, which implies $\sqrt{n}(\Delta_{HT}^* - \hat{\Delta}_{HT}) \xrightarrow{p} 0$. This result implies $\sigma^2 = \sigma_o^2$.

(ii) Proof For Ratio Estimator Δ_{Ratio},

We have

$$
\begin{aligned}
&\sqrt{n}(\Delta^* - \hat{\Delta})\\
&= \sqrt{n}\left(\sum_{i=1}^{n}\frac{D_i}{e_i^*}\right)^{-1}(\sum_{i=1}^{n}\frac{D_iY_i}{e_i^*}) - \sqrt{n}\left(\sum_{i=1}^{n}\frac{D_i}{\hat{e}_i}\right)^{-1}\left(\sum_{i=1}^{n}\frac{D_iY_i}{\hat{e}_i}\right)\\
&- \sqrt{n}\left(\sum_{i=1}^{n}\frac{1-D_i}{1-e_i^*}\right)^{-1}\left(\sum_{i=1}^{n}\frac{(1-D_i)Y_i}{1-e_i^*}\right) + \sqrt{n}\left(\sum_{i=1}^{n}\frac{1-D_i}{1-\hat{e}_i}\right)^{-1}\left(\sum_{i=1}^{n}\frac{1-D_iY_i}{1-\hat{e}_i}\right)\\
&= B_1^{(n)} + B_2^{(n)} + B_3^{(n)} + B_4^{(n)},
\end{aligned}
$$

where

$$
\begin{aligned}
B_1^{(n)} &= \sqrt{n}\left(\sum_{i=1}^{n}\frac{D_i}{e_i^*}\right)^{-1}\sum_{i=1}^{n}\left(\frac{D_iY_i}{e_i^*}-\frac{D_iY_i}{\hat{e}_i}\right),\\
B_2^{(n)} &= \sqrt{n}\left(\sum_{i=1}^{n}\frac{D_iY_i}{\hat{e}_i}\right)\left\{\left(\sum_{i=1}^{n}\frac{D_i}{e_i^*}\right)^{-1}-\left(\sum_{i=1}^{n}\frac{D_i}{\hat{e}_i}\right)^{-1}\right\},\\
B_3^{(n)} &= \sqrt{n}\left(\sum_{i=1}^{n}\frac{1-D_i}{1-e_i^*}\right)^{-1}\sum_{i=1}^{n}\left(\frac{(1-D_i)Y_i}{1-\hat{e}_i}-\frac{(1-D_i)Y_i}{1-e_i^*}\right),\\
B_4^{(n)} &= \sqrt{n}\left(\sum_{i=1}^{n}\frac{(1-D_i)Y_i}{1-\hat{e}_i}\right)\left\{\left(\sum_{i=1}^{n}\frac{1-D_i}{1-\hat{e}_i}\right)^{-1}-\left(\sum_{i=1}^{n}\frac{1-D_i}{1-e_i^*}\right)^{-1}\right\}.
\end{aligned}
$$

We only show that $B_1^{(n)}, B_2^{(n)} \xrightarrow{p} 0$, the proof for $B_3^{(n)}, B_4^{(n)} \xrightarrow{p} 0$ is similar so we simply omit it. We handle $B_1^{(n)}$ first. We note that

$$
\begin{aligned}
B_1^{(n)} &= \left(\frac{1}{n}\sum_{i}^{n}\frac{D_i}{e_i^*}\right)^{-1}\cdot\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\left(\frac{D_iY_i}{e_i^*}-\frac{D_iY_i}{\hat{e}_i}\right),\\
B_2^{(n)} &= \left(\frac{1}{n}\sum_{i=1}^{n}\frac{D_iY_i}{\hat{e}_i}\right)\cdot\left(\frac{1}{n}\sum_{i}^{n}\frac{D_i}{e_i^*}\right)^{-1}\left(\frac{1}{n}\sum_{i}^{n}\frac{D_i}{\hat{e}_i}\right)^{-1}\cdot\frac{\sqrt{n}}{n}\left(\sum_{i}^{n}\frac{D_i}{\hat{e}_i}-\frac{D_i}{e_i^*}\right).
\end{aligned}
$$

Firstly, from proof of (i), we know that

$$
\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\left(\frac{D_iY_i}{e_i^*}-\frac{D_iY_i}{\hat{e}_i}\right)\xrightarrow{p}0,\quad \frac{\sqrt{n}}{n}\sum_{i=1}^{n}\left(\frac{D_i}{e_i^*}-\frac{D_i}{\hat{e}_i}\right)\xrightarrow{p}0. \tag{63}
$$

We will use the same technique in the proof of (i) to show that

$$
\frac{1}{n}\sum_{i}^{n}\frac{D_i}{e_i^*}\xrightarrow{p}E\left(\frac{D}{e(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0)}\right)=1,\quad \frac{1}{n}\sum_{i}^{n}\frac{D_i}{\hat{e}_i}\xrightarrow{p}E\left(\frac{D}{e(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0)}\right)=1, \tag{64}
$$

$$
\frac{1}{n}\sum_{i}^{n}\frac{D_iY_i}{\hat{e}_i}\xrightarrow{p}E\left(\frac{DY}{e(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0)}\right)=E\{Y(1)\}. \tag{65}
$$

With (63), (64) and (65), by the Continuous mapping theorem, we can conclude that $B_1^{(n)}, B_2^{(n)} \xrightarrow{p}$

0. And analogously we have $B_3^{(n)}, B_4^{(n)} \xrightarrow{p} 0$. So $\sqrt{n}(\Delta^*_{Ratio} - \hat{\Delta}^*_{Ratio}) \xrightarrow{p} 0$, which implies $\sigma^2 = \sigma_o^2$.

We define $L(\boldsymbol{\beta}) = (1/n)\sum_{i=1}^n D_i/e(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta})$, $D(\boldsymbol{\beta}) = L(\boldsymbol{\beta}) - L(\boldsymbol{\beta}_0)$. Assume $\boldsymbol{\beta}$ is a consistent estimator of $\boldsymbol{\beta}_0$, we have

$$D(\boldsymbol{\beta}) = \left\{\frac{\partial L(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}_0} + o_p(1)\right\}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^{\mathrm{T}}$$

So if $\boldsymbol{\beta} \xrightarrow{p} \boldsymbol{\beta}_0$, we have $D(\boldsymbol{\beta}) \xrightarrow{p} 0$. So we have $D(\boldsymbol{\beta}^*) \xrightarrow{p} 0$ and $D(\hat{\boldsymbol{\beta}}) \xrightarrow{p} 0$, which imply $(1/n)\sum_{i=1}^n D_i/e(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}^*) = L(\boldsymbol{\beta}^*) = L(\boldsymbol{\beta}_0) + D(\boldsymbol{\beta}^*) \xrightarrow{p} E\{D/e(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0)\} = 1$ and $(1/n)\sum_{i=1}^n D_i/e(\boldsymbol{X}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}) = L(\hat{\boldsymbol{\beta}}) = L(\boldsymbol{\beta}_0) + D(\hat{\boldsymbol{\beta}}) \xrightarrow{p} E\{D/e(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0)\} = 1$.

So far, we have showed proved (64). And we can obtain (65) by a similar argument.