A Simple and Efficient Estimation Method for Models with Non-ignorable Missing Data

Chunrong Ai

Department of Economics, University of Florida, U.S.

Oliver Linton

Faculty of Economics, University of Cambridge, U.K.

Zheng Zhang

Institute of Statistics and Big Data, Renmin University of China, China

Supplementary Material

S1 Assumptions

We first introduce the smoothness classes of functions used in the nonparametric estimation; see e.g. Stone (1982, 1994), Robinson (1988), Newey (1997), Horowitz (2012) and Chen (2007). Suppose that $\mathcal{X}$ is the Cartesian product of r-compact intervals. Let $0 < \delta \leq 1$. A fucntion f on $\mathcal{X}$ is said to satisfy a Hölder condition with exponent δ if there is a positive constant L usch that $\|f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2)\| \leq L\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^{\delta}$ for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$. Given a

r-tuple $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_r)$ of nonnegative integer, denote $[\boldsymbol{\alpha}] = \alpha_1 + \cdots + \alpha_r$ and let $D^{\boldsymbol{\alpha}}$ denote the differential operator defined by $D^{\boldsymbol{\alpha}} = \frac{\partial^{[\boldsymbol{\alpha}]}}{\partial x_1^{\alpha_1} \cdots \partial x_r^{\alpha_r}}$, where $\boldsymbol{x} = (x_1, ..., x_r)$.

Definition 1. *Let s be a nonnegative integer and $s := s_0 + \delta$. The function f on $\mathcal{X}$ is said to be s-*smooth *if it is s_0 times continuously differentiable on $\mathcal{X}$ and $D^{\boldsymbol{\alpha}} f$ satisfies a Hölder condition with exponent δ for all $\boldsymbol{\alpha}$ with $[\boldsymbol{\alpha}] = s_0$.*

We use the notation $a^{\otimes 2} := aa^{\intercal}$ for a vector a. The following notation are needed for our proof:

$$O(\boldsymbol{Z}) := \frac{1 - \pi(\boldsymbol{Z}; \gamma_0)}{\pi(\boldsymbol{Z}; \gamma_0)}, \quad \boldsymbol{S}_0(\boldsymbol{Z}) := -\frac{\nabla_\gamma \pi(\boldsymbol{Z}; \gamma_0)}{1 - \pi(\boldsymbol{Z}; \gamma_0)} \ , \tag{S1.1}$$

$$m(\boldsymbol{X}) := \frac{\mathbb{E}[O(\boldsymbol{Z})\boldsymbol{S}_0(\boldsymbol{Z})|\boldsymbol{X}]}{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}, \quad R(\boldsymbol{X}) := \frac{\mathbb{E}[O(\boldsymbol{Z})U(\boldsymbol{Z})|\boldsymbol{X}]}{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]} \ , \tag{S1.2}$$

$$\boldsymbol{S}_1(T, \boldsymbol{Z}; \gamma_0) := \left(1 - \frac{T}{\pi(\boldsymbol{Z}; \gamma_0)}\right) m(\boldsymbol{X}) \ , \tag{S1.3}$$

$$S_2(T, \boldsymbol{Z}; \gamma_0, \theta_0) := -\frac{T}{\pi(\boldsymbol{Z}; \gamma_0)} U(\boldsymbol{Z}) + \theta_0 - \left(1 - \frac{T}{\pi(\boldsymbol{Z}; \gamma_0)}\right) R(\boldsymbol{X}) \ , \tag{S1.4}$$

$$\kappa^{\intercal} := \mathbb{E}\left[\frac{\nabla_\gamma \pi(\boldsymbol{Z}; \gamma_0)^{\intercal}}{\pi(\boldsymbol{Z}; \gamma_0)} \{R(\boldsymbol{Z}) - U(\boldsymbol{X})\}\right] \cdot \mathbb{E}\left[\frac{m(\boldsymbol{X})}{\pi(\boldsymbol{Z}; \gamma_0)} \nabla_\gamma \pi(\boldsymbol{Z}; \gamma_0)^{\intercal}\right]^{-1} \ , \tag{S1.5}$$

$$\boldsymbol{V}_{\gamma_0} = \mathbb{E}\left[\frac{1 - \pi(\boldsymbol{Z}; \gamma_0)}{\pi(\boldsymbol{Z}; \gamma_0)} m(\boldsymbol{X})^{\otimes 2}\right]^{-1} \text{ (the efficient variance bound of } \gamma_0) \ , \tag{S1.6}$$

$$V_{\theta_0} = Var\left(S_2(T, \boldsymbol{Z}; \gamma_0, \theta_0) - \kappa^{\intercal} \boldsymbol{S}_1(T, \boldsymbol{Z}; \gamma_0)\right)$$

$$(\text{the efficient variance bound of } \theta_0) \ . \tag{S1.7}$$

The following assumptions are maintained in this paper:

Assumption 1. There exists a nonresponse instrumental variable $\mathbf{X_2}$, i.e., $\mathbf{X} = (\mathbf{X}_1^\intercal, \mathbf{X}_2^\intercal)^\intercal$, such that $\boldsymbol{X}_2$ is independent of T given $\boldsymbol{X}_1$ and Y; furthermore, $\boldsymbol{X}_2$ is correlated with Y.

Assumption 2. The support of $\mathbf{X}$, which is denoted by $\mathcal{X}$, is a Cartesian product of r-compact intervals, and we denote $\boldsymbol{X} = (X_1, ..., X_r)^\top$.

Assumption 3. The functions $\mathbb{E}[O(\boldsymbol{Z})S_0(\boldsymbol{Z})|\mathbf{X} = \boldsymbol{x}]$, $\mathbb{E}[O(\boldsymbol{Z})U(\boldsymbol{Z})|\mathbf{X} = \boldsymbol{x}]$ and $\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X} = \boldsymbol{x}]$ are s-smooth in $\boldsymbol{x}$, where $s > 0$.

Assumption 4. There exist two finite positive constants $\underline{a}$ and $\overline{a}$ such that the smallest (resp. largest) eigenvalue of $\mathbb{E}[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})]$ is bounded away from $\underline{a}$ (resp. $\overline{a}$) uniformly in K, i.e.,

$$0 < \underline{a} \leq \lambda_{\min}(\mathbb{E}[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top]) \leq \lambda_{\max}(\mathbb{E}[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top]) \leq \overline{a} < \infty \ .$$

Remark 1. Asssumption 4 implies that following results:

1.
$$\mathbb{E}[\|u_K(\boldsymbol{X})\|^2] = \operatorname{tr}\left(\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]\right) = O(K) \ ; \tag{S1.8}$$

2. the matrices $\bar{a} \cdot I_{K\times K} - \mathbb{E}[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top]$ and $\mathbb{E}[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top] - \underline{a} \cdot$

$I_{K\times K}$ are positive definite, and

$$\underline{a} \le \inf_{k\in\{1,\ldots,K\}} \mathbb{E}[u_{kK}(\boldsymbol{X})^2] \le \sup_{k\in\{1,\ldots,K\}} \mathbb{E}[u_{kK}(\boldsymbol{X})^2] \le \overline{a} \ . \tag{S1.9}$$

Assumption 5. (i) The parameter spaces Γ and Θ are compact; (ii) The efficient score function $\boldsymbol{S}_{eff}(T,\boldsymbol{Z};\gamma,\theta) := (\boldsymbol{S}_1^\intercal(T,\boldsymbol{Z};\gamma), S_2(T,\boldsymbol{Z};\gamma,\theta))^\intercal$ is continuously differentiable at each $(\gamma,\theta)\in\Gamma\times\Theta$, and $\mathbb{E}\left[\partial\boldsymbol{S}_{eff}(\gamma,\theta)/\partial(\gamma^\top,\theta)\right]$ is nonsingular at (γ_0,θ_0).

Assumption 6. The response probability $\pi(\boldsymbol{x},y;\gamma)$ satisfies the following conditions:

1. there exist two positive constants $\bar{c}$ and $\underline{c}$ such that $0<\underline{c}\le\pi(\boldsymbol{x},y;\gamma)\le\bar{c}<1$ for all $\gamma\in\Gamma$ and $(\boldsymbol{x},y)\in\mathcal{X}\times\mathbb{R}$;

2. $\pi(\boldsymbol{x},y;\gamma)$ is twice continuously differentiable in $\gamma\in\Gamma$, and the derivatives are uniformly bounded.

Assumption 7. Suppose $K\to\infty$ and $K^3/N\to 0$.

S2 Some useful results

We present some results which will be used in the proof of Theorems 1 and 2.

S2.1 Matrix inversion formula

- (General Formula) Let $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{C}^{-1}+\mathbf{D}\mathbf{A}^{-1}\mathbf{B}$ be non-singular square matrices; then

$$(\mathbf{A}+\mathbf{BCD})^{-\mathbf{1}} = \mathbf{A}^{-\mathbf{1}} - \mathbf{A}^{-\mathbf{1}}\mathbf{B}(\mathbf{C}^{-\mathbf{1}} + \mathbf{D}\mathbf{A}^{-\mathbf{1}}\mathbf{B})^{-\mathbf{1}}\mathbf{D}\mathbf{A}^{-\mathbf{1}} \ . \qquad \text{(S2.10)}$$

- (Matrix Inversion in Block form) Let a $(m+1)\times(m+1)$ matrix $\boldsymbol{M}$ be partitioned into a block form

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{A}\ , & \boldsymbol{b} \\ \boldsymbol{b}^\top\ , & d \end{bmatrix}$$

where $\boldsymbol{A}$ is a $m\times m$ matrix, $\boldsymbol{b}$ is a m dimensional column vector, d is a constant. Then

$$\boldsymbol{M}^{-1} = \begin{bmatrix} \boldsymbol{A}^{-1} + \frac{1}{k}\boldsymbol{A}^{-1}\boldsymbol{b}\boldsymbol{b}^\top\boldsymbol{A}^{-1}, & -\frac{1}{k}\boldsymbol{A}^{-1}\boldsymbol{b} \\ -\frac{1}{k}\boldsymbol{b}^\top\boldsymbol{A}^{-1}, & \frac{1}{k} \end{bmatrix}, \qquad \text{(S2.11)}$$

where $k = d - \boldsymbol{b}^\top\boldsymbol{A}^{-1}\boldsymbol{b}$.

S2.2 Discussion on u_K

To construct the GMM estimator, we need to specify the matching function $u_K(\boldsymbol{X})$ The most common class of functions are power series. Suppose the dimension of covariate $\boldsymbol{X}$ is $r\in\mathbb{N}$, namely $\boldsymbol{X} = (X_1, ..., X_r)^\top$. Let $\lambda = (\lambda_1, \ldots \lambda_r)^\top$ be an r-dimensional vector of nonnegative integers (multi-indices), with norm $|\lambda| = \sum_{j=1}^{r} \lambda_j$. Let $(\lambda(k))_{k=1}^{\infty}$ be a sequence that includes all distinct multi-indicesand satisfies $|\lambda(k)| \leq |\lambda(k+1)|$, and let $\boldsymbol{X}^{\lambda} = \prod_{j=1}^{r} X_j^{\lambda_j}$. For a sequence $\lambda(k)$ we consider the series $u_{kK}(\boldsymbol{X}) = \boldsymbol{X}^{\lambda(k)}$, $k \in \{1, ..., K\}$. Newey (1997) showed the following property for the power series: there exists a universal constant $C > 0$ such that

$$\zeta(K) := \sup_{\boldsymbol{x} \in \mathcal{X}} \|u_K(\boldsymbol{x})\| \leq CK \ , \tag{S2.12}$$

where $\|\cdot\|$ denotes the usual matrix norm $\|A\| = \sqrt{\mathrm{tr}(A^{\top}A)}$.

S2.3 Convergence rate of L^2 approximation

Suppose $f : \mathbb{R}^r \to \mathbb{R}$ is the function we want to approximate. Let $f_K(\mathbf{X})$ be the L^2-projection of $f(\boldsymbol{X})$ on the space linearly spanned by $u_K(\boldsymbol{X})$, i.e.

$$f_K(\mathbf{X}) = \beta_K^{\top} u_K(\boldsymbol{X}) \tag{S2.13}$$

where

$$\beta_K := \mathbb{E}\left[u_K(\boldsymbol{X}) u_K(\boldsymbol{X})^{\top}\right]^{-1} \mathbb{E}\left[u_K(\boldsymbol{X}) f(\boldsymbol{X})\right] \ .$$

In this section, we establish the L^2-convergence rate of $f_K(\boldsymbol{X})$ to $f(\boldsymbol{X})$, which will be used for proving the theorems of our paper.

Lemma 1. *Under Assumpitons 2 and 4, suppose the function $f : \mathbb{R}^r \to \mathbb{R}$*

is s-smooth and f_K is defined by (S2.13)*, then we have*

$$\mathbb{E}\left[|f(\boldsymbol{X}) - f_K(\boldsymbol{X})|^2\right] = O\left(K^{-\frac{2s}{r}}\right) \ .$$

Proof. Since $f(\boldsymbol{x})$ is s-smooth and the support $\mathcal{X}$ is compact by Assumpiton 2, from Section 2.3.1 of Chen (2007), we know that there exists $\beta^* \in \mathbb{R}^K$ such that

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \left|f(\boldsymbol{x}) - (\beta^*)^\top u_K(\boldsymbol{x})\right| = O(K^{-\frac{s}{r}}) \ .$$

We first claim that

$$\|\beta_K - \beta^*\| = O(K^{-\frac{s}{r}}) \ . \tag{S2.14}$$

With the claim (S2.14), Cauchy-Schwarz inequality, and Assumption 4, we can obtain that

$$\begin{aligned}
&\mathbb{E}\left[|f(\boldsymbol{X}) - f_K(\boldsymbol{X})|^2\right] \\
=&\int_{\mathcal{X}} \left\{(\beta_K - \beta^*)^\top u_K(\boldsymbol{x}) + \left[(\beta^*)^\top u_K(\boldsymbol{x}) - f(\boldsymbol{x})\right]\right\}^2 dF_X(\boldsymbol{x}) \\
\leq&2(\beta_K - \beta^*)^\top \int_{\mathcal{X}} u_K(\boldsymbol{x}) u_K(\boldsymbol{x})^\top dF_X(\boldsymbol{x})(\beta_K - \beta^*) + 2\int_{\mathcal{X}} \left[(\beta^*)^\top u_K(\boldsymbol{x}) - f(\boldsymbol{x})\right]^2 dF_X(\boldsymbol{x}) \\
\leq&2\|\beta_K - \beta^*\|^2 \cdot \lambda_{\max}\left(\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]\right) + 2\sup_{\boldsymbol{x}\in\mathcal{X}} \left\|f(\boldsymbol{x}) - (\beta^*)^\top u_K(\boldsymbol{x})\right\|^2 \\
=&2\|\beta_K - \beta^*\|^2 \cdot O(1) + O(K^{-\frac{2s}{r}}) = O(K^{-\frac{2s}{r}}) \ .
\end{aligned}$$

We now prove the claim (S2.14). Note that

$$\beta_K - \beta^* = \mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1} \mathbb{E}\left[u_K(\boldsymbol{X})f(\boldsymbol{X})^\top\right] - \beta^*$$

$$
\begin{aligned}
=&\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\mathbb{E}\left[u_K(\boldsymbol{X})f(\boldsymbol{X})^\top\right]\\
&-\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\beta^*\right]\\
=&\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\mathbb{E}\left[u_K(\boldsymbol{X})\left\{f(\boldsymbol{X})-(\beta^*)^\top u_K(\boldsymbol{X})\right\}\right]\ .
\end{aligned}
$$

Then

$$
\begin{aligned}
&\|\beta_K-\beta^*\|^2\\
=&\mathrm{tr}\left(\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\mathbb{E}\left[u_K(\boldsymbol{X})\left\{f(\boldsymbol{X})-(\beta^*)^\top u_K(\boldsymbol{X})\right\}\right]^{\otimes 2}\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\right)\\
\leq&\lambda_{\max}\left(\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\right)\\
&\cdot\,\mathrm{tr}\left(\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-\frac{1}{2}}\mathbb{E}\left[u_K(\boldsymbol{X})\left\{f(\boldsymbol{X})-(\beta^*)^\top u(\boldsymbol{X})\right\}\right]^{\otimes 2}\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-\frac{1}{2}}\right)\\
=&\lambda_{\max}\left(\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\right)\\
&\cdot\,\mathbb{E}\left[\left\{f(\boldsymbol{X})-(\beta^*)^\top u_K(\boldsymbol{X})\right\}u_K(\boldsymbol{X})^\top\right]\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\mathbb{E}\left[u_K(\boldsymbol{X})\left\{f(\boldsymbol{X})-(\beta^*)^\top u(\boldsymbol{X})\right\}\right]\\
=&\lambda_{\max}\left(\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\right)\cdot\mathbb{E}\left[\left|\mathbb{E}\left[\left\{f(\boldsymbol{X})-(\beta^*)^\top u_K(\boldsymbol{X})\right\}u_K(\boldsymbol{X})^\top\right]\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}u_K(\boldsymbol{X})\right|^2\right]\\
\leq&\lambda_{\max}\left(\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\right)\mathbb{E}\left[\left|f(\boldsymbol{X})-(\beta^*)^\top u_K(\boldsymbol{X})\right|^2\right]\\
\leq&\lambda_{\max}\left(\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\right)\sup_{\boldsymbol{x}\in\mathcal{X}}\left|f(\boldsymbol{x})-(\beta^*)^\top u_K(\boldsymbol{x})\right|^2=O(K^{-\frac{2s}{r}})\ ,
\end{aligned}
$$

where the first inequality follow from the fact that $\mathrm{tr}(AB)\leq\lambda_{\max}(B)\mathrm{tr}(A)$ for any symmetric matrix B and positive semidefinite matrix A; the second inequality follows from the fact that

$$
\mathbb{E}\left[\left\{f(\boldsymbol{X})-(\beta^*)^\top u_K(\boldsymbol{X})\right\}u_K(\boldsymbol{X})^\top\right]\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}u_K(\boldsymbol{X})
$$

is the L^2-projection of $f(\boldsymbol{X})-(\beta^*)^\top u_K(\boldsymbol{X})$ on the space spanned by $u_K(\mathbf{X})$, which implies

$$
\left\|\mathbb{E}\left[\left\{f(\boldsymbol{X})-(\beta^*)^\top u_K(\boldsymbol{X})\right\}u_K(\boldsymbol{X})^\top\right]\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}u_K(\boldsymbol{X})\right\|_{L^2}\leq\left\|f(\boldsymbol{X})-(\beta^*)^\top u_K(\boldsymbol{X})\right\|_{L^2}\ .
$$

This complete the proof of the lemma.

□

S3 Proof of Theorem 1

Define the objective functions:

$$\hat{Q}_N(\gamma,\theta) := \left\{ \frac{1}{N}\sum_{i=1}^{N} g_K(T_i, \boldsymbol{Z}_i; \gamma, \theta) \right\}^{\top} \widehat{\boldsymbol{W}}_0^{-1} \left\{ \frac{1}{N}\sum_{i=1}^{N} g_K(T_i, \boldsymbol{Z}_i; \gamma, \theta) \right\} ,$$

and

$$Q_0(\gamma,\theta) := \mathbb{E}\left[g_K(T,\boldsymbol{Z};\gamma,\theta)\right]^{\top} \mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1} \mathbb{E}\left[g_K(T,\boldsymbol{Z};\gamma,\theta)\right].$$

By definition, $(\check{\gamma},\check{\theta})$ and (γ_0,θ_0) are unique minimizers of $\hat{Q}_N(\cdot,\cdot)$ and $Q_0(\cdot,\cdot)$ respectively. Note that

$$
\begin{aligned}
&|\hat{Q}_N(\gamma,\theta) - Q_0(\gamma,\theta)| \\
\le& \left| \left\{ \frac{1}{N}\sum_{i=1}^{N} g_K(T_i,\boldsymbol{Z}_i;\gamma,\theta) - \mathbb{E}\left[g_K(T,\boldsymbol{Z};\gamma,\theta)\right] \right\}^{\top} \widehat{\boldsymbol{W}}_0^{-1} \left\{ \frac{1}{N}\sum_{i=1}^{N} g_K(T_i,\boldsymbol{Z}_i;\gamma,\theta) \right\} \right| && \text{(S3.15)} \\
&+ \left| \mathbb{E}\left[g_K(T,\boldsymbol{Z};\gamma,\theta)\right]^{\top} \left\{ \widehat{\boldsymbol{W}}_0^{-1} - \mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1} \right\} \left\{ \frac{1}{N}\sum_{i=1}^{N} g_K(T_i,\boldsymbol{Z}_i;\gamma,\theta) \right\} \right| && \text{(S3.16)} \\
&+ \left| \mathbb{E}\left[g_K(T,\boldsymbol{Z};\gamma,\theta)\right]^{\top} \mathbb{E}[\widehat{\boldsymbol{W}}]_0^{-1} \left\{ \frac{1}{N}\sum_{i=1}^{N} g_K(T_i,\boldsymbol{Z}_i;\gamma,\theta) - \mathbb{E}\left[g_K(T,\boldsymbol{Z};\gamma,\theta)\right] \right\} \right|. && \text{(S3.17)}
\end{aligned}
$$

Consider the term (S3.15). Let

$$\boldsymbol{a}_1 := \frac{1}{N}\sum_{i=1}^{N} g_K(T_i,\boldsymbol{Z}_i;\gamma,\theta) - \mathbb{E}\left[g_K(T,\boldsymbol{Z};\gamma,\theta)\right] \text{ and } \boldsymbol{a}_2 := \frac{1}{N}\sum_{i=1}^{N} g_K(T_i,\boldsymbol{Z}_i;\gamma,\theta),$$

and $\lambda_{\max}(A)$ (resp. $\lambda_{\min}(A)$) denote the maximum (resp. minimum) eigenvalue of a matrix A. We have that

$$|(\text{S3.15})|^2 = \boldsymbol{a}_1^{\top} \widehat{\boldsymbol{W}}_0^{-1} \boldsymbol{a}_2 \boldsymbol{a}_2^{\top} \widehat{\boldsymbol{W}}_0^{-1} \boldsymbol{a}_1 \le \lambda_{\max}\left(\boldsymbol{a}_2\boldsymbol{a}_2^{\top}\right) \cdot \boldsymbol{a}_1^{\top} \widehat{\boldsymbol{W}}_0^{-1} \widehat{\boldsymbol{W}}_0^{-1} \boldsymbol{a}_1$$

$$\leq \lambda_{\max}\left(\boldsymbol{a}_2\boldsymbol{a}_2^\top\right)\cdot\lambda_{\max}\left(\widehat{\boldsymbol{W}}_0^{-2}\right)\cdot\|\boldsymbol{a}_1\|^2.$$

By Assumptions 4 and 6.1, we have $\lambda_{\max}\left(\boldsymbol{a}_2\boldsymbol{a}_2^\top\right)=O_p(1)$ and $\lambda_{\max}\left(\widehat{\boldsymbol{W}}_0^{-2}\right)=\left[\lambda_{\min}\left(\widehat{\boldsymbol{W}}_0\right)\right]^{-2}=O_p(1)$. Note that

$$\begin{aligned}\mathbb{E}\left[\|\boldsymbol{a}_1\|^2\right]=&\frac{1}{N}\mathbb{E}\left[\|g_K(T,\boldsymbol{Z};\gamma,\theta)-\mathbb{E}\left[g_K(T,\boldsymbol{Z};\gamma,\theta)\right]\|^2\right]\leq\frac{1}{N}\mathbb{E}\left[\|g_K(T,\boldsymbol{Z};\gamma,\theta)\|^2\right]\\=&\frac{1}{N}\mathbb{E}\left[\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma)}\right)^2\cdot\|u_K(\boldsymbol{X})\|^2\right]+\frac{1}{N}\mathbb{E}\left[\left(\theta-\frac{T}{\pi(\boldsymbol{Z};\gamma)}U(\boldsymbol{Z})\right)^2\right]\\\leq&\frac{1}{N}\cdot O(1)\cdot\mathbb{E}[\|u_K(\boldsymbol{X})\|^2]+\frac{1}{N}\cdot O(1)\leq O\left(\frac{K}{N}\right),\end{aligned}$$

where the second inequality holds because of Assumption 6.1 that $\pi(\boldsymbol{Z};\gamma)$ is uniformly bounded away from zero; the last inequality holds because

$$\mathbb{E}[\|u_K(\boldsymbol{X})\|^2]=\mathrm{tr}\left(\mathbb{E}[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top]\right)\leq\lambda_{\max}\left(\mathbb{E}[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top]\right)\cdot K=O(K).$$

Therefore, (S3.15) is of $O_p(\sqrt{K/N})$ by Chebyshev's inequality. Similarly, (S3.17) is also of $O_p(\sqrt{K/N})$.

We next consider (S3.16). Note that

$$\begin{aligned}|(\mathrm{S3.16})|^2=&\left|\mathbb{E}\left[\boldsymbol{a}_2\right]^\top\left\{\widehat{\boldsymbol{W}}_0^{-1}-\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\right\}\boldsymbol{a}_2\right|^2=\mathbb{E}\left[\boldsymbol{a}_2\right]^\top\left\{\widehat{\boldsymbol{W}}_0^{-1}-\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\right\}\boldsymbol{a}_2\boldsymbol{a}_2^\top\left\{\widehat{\boldsymbol{W}}_0^{-1}-\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\right\}\mathbb{E}\left[\boldsymbol{a}_2\right]\\\leq&\lambda_{\max}\left(\boldsymbol{a}_2\boldsymbol{a}_2^\top\right)\cdot\mathbb{E}\left[\boldsymbol{a}_2\right]^\top\left\{\widehat{\boldsymbol{W}}_0^{-1}-\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\right\}\left\{\widehat{\boldsymbol{W}}_0^{-1}-\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\right\}\mathbb{E}\left[\boldsymbol{a}_2\right]\\=&\lambda_{\max}\left(\boldsymbol{a}_2\boldsymbol{a}_2^\top\right)\cdot\mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{a}_2\right]\mathbb{E}\left[\boldsymbol{a}_2\right]^\top\left\{\widehat{\boldsymbol{W}}_0^{-1}-\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\right\}\left\{\widehat{\boldsymbol{W}}_0^{-1}-\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\right\}\right)\\\leq&\lambda_{\max}\left(\boldsymbol{a}_2\boldsymbol{a}_2^\top\right)\cdot\lambda_{\max}\left(\mathbb{E}\left[\boldsymbol{a}_2\right]\mathbb{E}\left[\boldsymbol{a}_2\right]^\top\right)\cdot\mathrm{tr}\left(\left\{\widehat{\boldsymbol{W}}_0^{-1}-\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\right\}\left\{\widehat{\boldsymbol{W}}_0^{-1}-\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\right\}\right),\end{aligned}$$

where the last inequality follows from the fact that $\mathrm{tr}(AB)\leq\lambda_{\max}(B)\mathrm{tr}(A)$ for any symmetric matrix B and positive semidefinite matrix A. By Assumptions 4 and 6.1, we have that $\lambda_{\max}\left(\boldsymbol{a}_2\boldsymbol{a}_2^\top\right)=O_p(1)$ and $\lambda_{\max}\left(\mathbb{E}\left[\boldsymbol{a}_2\right]\mathbb{E}\left[\boldsymbol{a}_2\right]^\top\right)=O(1)$. Note that

$$
\begin{aligned}
&\operatorname{tr}\left(\left\{\widehat{\boldsymbol{W}}_0^{-1}-\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\right\}\left\{\widehat{\boldsymbol{W}}_0^{-1}-\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\right\}\right)\\
=&\operatorname{tr}\left(\widehat{\boldsymbol{W}}_0^{-1}\left\{\mathbb{E}[\widehat{\boldsymbol{W}}_0]-\widehat{\boldsymbol{W}}_0\right\}\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\left\{\mathbb{E}[\widehat{\boldsymbol{W}}_0]-\widehat{\boldsymbol{W}}_0\right\}\widehat{\boldsymbol{W}}_0^{-1}\right)\\
=&\operatorname{tr}\left(\widehat{\boldsymbol{W}}_0^{-1}\widehat{\boldsymbol{W}}_0^{-1}\left\{\mathbb{E}[\widehat{\boldsymbol{W}}_0]-\widehat{\boldsymbol{W}}_0\right\}\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\left\{\mathbb{E}[\widehat{\boldsymbol{W}}_0]-\widehat{\boldsymbol{W}}_0\right\}\right)\\
\le&\lambda_{\max}\left(\widehat{\boldsymbol{W}}_0^{-1}\widehat{\boldsymbol{W}}_0^{-1}\right)\cdot\operatorname{tr}\left(\left\{\mathbb{E}[\widehat{\boldsymbol{W}}_0]-\widehat{\boldsymbol{W}}_0\right\}\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\left\{\mathbb{E}[\widehat{\boldsymbol{W}}_0]-\widehat{\boldsymbol{W}}_0\right\}\right)\\
=&\lambda_{\max}\left(\widehat{\boldsymbol{W}}_0^{-1}\widehat{\boldsymbol{W}}_0^{-1}\right)\cdot\operatorname{tr}\left(\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\left\{\mathbb{E}[\widehat{\boldsymbol{W}}_0]-\widehat{\boldsymbol{W}}_0\right\}\left\{\mathbb{E}[\widehat{\boldsymbol{W}}_0]-\widehat{\boldsymbol{W}}_0\right\}\right)\\
\le&\lambda_{\max}\left(\widehat{\boldsymbol{W}}_0^{-1}\widehat{\boldsymbol{W}}_0^{-1}\right)\cdot\lambda_{\max}\left(\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\right)\operatorname{tr}\left(\left\{\mathbb{E}[\widehat{\boldsymbol{W}}_0]-\widehat{\boldsymbol{W}}_0\right\}\left\{\mathbb{E}[\widehat{\boldsymbol{W}}_0]-\widehat{\boldsymbol{W}}_0\right\}\right)\\
\le&O_p(1)\cdot O(1)\cdot\left\|\widehat{\boldsymbol{W}}_0-\mathbb{E}[\widehat{\boldsymbol{W}}_0]\right\|^2=O_p\left(\frac{K^3}{N}\right),
\end{aligned}
$$

where the last inequality follows from the following fact and a use of Chebyshev's inequality:

$$
\begin{aligned}
\mathbb{E}\left[\left\|\widehat{\boldsymbol{W}}_0-\mathbb{E}[\widehat{\boldsymbol{W}}_0]\right\|^2\right]=&\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top-\mathbb{E}[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top]\right\|^2\right]\\
=&\frac{1}{N}\mathbb{E}\left[\left\|u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top-\mathbb{E}[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top]\right\|^2\right]\\
\le&\frac{1}{N}\mathbb{E}[\|u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\|]\\
\le&\frac{1}{N}\sup_{\boldsymbol{x}\in\mathcal{X}}\|u_K(\boldsymbol{x})\|^2\cdot\mathbb{E}[\|u_K(\boldsymbol{X})\|^2]=\frac{1}{N}\cdot O(K^2)\cdot O(K)\\
=&\frac{K^3}{N}.
\end{aligned}
$$

Then (S3.16) is of $O_p(\sqrt{K^3/N})$. Therefore, for each $(\gamma,\theta)\in\Gamma\times\Theta$, we have

$$
\begin{aligned}
|\hat{Q}_N(\gamma,\theta)-Q_0(\gamma,\theta)|\le|(\text{S3.15})|+|(\text{S3.16})|+|(\text{S3.17})|=&O_p\left(\sqrt{\frac{K}{N}}\right)+O_p\left(\sqrt{\frac{K^3}{N}}\right)+O_p\left(\sqrt{\frac{K}{N}}\right)\\
=&o_p(1). \qquad \text{(S3.18)}
\end{aligned}
$$

where the last equality follows from Assumption 7. Next, we strengthen above convergence result to $\sup_{(\gamma,\theta)\in\Gamma\times\Theta}|\hat{Q}_N(\gamma,\theta)-Q_0(\gamma,\theta)|=o_p(1)$. Note that

$$\hat{Q}_N(\gamma,\theta)=\left\{\frac{1}{N}\sum_{i=1}^N\left(1-\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)}\right)u_K(\boldsymbol{X}_i)\right\}^\top\left[\frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top\right]^{-1}\left\{\frac{1}{N}\sum_{i=1}^N\left(1-\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)}\right)u_K(\boldsymbol{X}_i)\right\}$$
$$+\left\{\frac{1}{N}\sum_{i=1}^N\left(\theta-\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)}U(\boldsymbol{Z}_i)\right)\right\}^2$$
$$=\frac{1}{N}\sum_{i=1}^N\left(\widehat{\mathbb{E}}\left[1-\frac{T}{\pi(\boldsymbol{Z};\gamma)}\middle|\boldsymbol{X}=\boldsymbol{X}_i\right]\right)^2+\left\{\frac{1}{N}\sum_{i=1}^N\left(\theta-\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)}U(\boldsymbol{Z}_i)\right)\right\}^2,$$

where the operator $\widehat{\mathbb{E}}[\cdot|\boldsymbol{X}]$ is defined by

$$\widehat{\mathbb{E}}\left[\phi(\boldsymbol{Z})|\boldsymbol{X}\right]:=\left\{\frac{1}{N}\sum_{l=1}^N\phi(\boldsymbol{Z}_i)u_K(\boldsymbol{X}_l)\right\}^\top\left[\frac{1}{N}\sum_{l=1}^N u_K(\boldsymbol{X}_l)u_K(\boldsymbol{X}_l)^\top\right]^{-1}u_K(\boldsymbol{X}),\ \forall\phi(\cdot)\in L^2,$$

which is the least square projection of $\phi(\boldsymbol{Z})$ on space linearly spanned by the basis $u_K(\boldsymbol{X})$. Note that

$$\begin{aligned}\|\nabla_{(\gamma,\theta)}\hat{Q}_N(\gamma,\theta)\|\leq&\frac{2}{N}\sum_{i=1}^N\left|\widehat{\mathbb{E}}\left[1-\frac{T}{\pi(\boldsymbol{Z};\gamma)}\middle|\boldsymbol{X}=\boldsymbol{X}_i\right]\right|\cdot\left\|\widehat{\mathbb{E}}\left[\frac{T}{\pi(\boldsymbol{Z};\gamma)^2}\nabla_\gamma\pi(\boldsymbol{Z};\gamma)\middle|\boldsymbol{X}=\boldsymbol{X}_i\right]\right\|\\&+2\cdot\left|\frac{1}{N}\sum_{i=1}^N\left(\theta-\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)}U(\boldsymbol{Z}_i)\right)\right|\cdot\left\|\frac{1}{N}\sum_{i=1}^N\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)^2}\nabla_\gamma\pi(\boldsymbol{Z};\gamma)U(\boldsymbol{Z}_i)\right\|\\&+2\cdot\left|\frac{1}{N}\sum_{i=1}^N\left(\theta-\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)}U(\boldsymbol{Z}_i)\right)\right|\\\leq&\frac{1}{N}\sum_{i=1}^N\left|\widehat{\mathbb{E}}\left[1-\frac{T}{\pi(\boldsymbol{Z};\gamma)}\middle|\boldsymbol{X}=\boldsymbol{X}_i\right]\right|^2+\frac{1}{N}\sum_{i=1}^N\left\|\widehat{\mathbb{E}}\left[\frac{T}{\pi(\boldsymbol{Z};\gamma)^2}\nabla_\gamma\pi(\boldsymbol{Z};\gamma)\middle|\boldsymbol{X}=\boldsymbol{X}_i\right]\right\|^2\\&+2\cdot\left|\frac{1}{N}\sum_{i=1}^N\left(\theta-\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)}U(\boldsymbol{Z}_i)\right)\right|\cdot\left\|\frac{1}{N}\sum_{i=1}^N\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)^2}\nabla_\gamma\pi(\boldsymbol{Z};\gamma)U(\boldsymbol{Z}_i)\right\|\\&+2\cdot\left|\frac{1}{N}\sum_{i=1}^N\left(\theta-\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)}U(\boldsymbol{Z}_i)\right)\right|.\end{aligned}$$

Using the least square projection property and Assumption 6, we have that for all (γ,θ):

$$\frac{1}{N}\sum_{i=1}^N\left|\widehat{\mathbb{E}}\left[1-\frac{T}{\pi(\boldsymbol{Z};\gamma)}\middle|\boldsymbol{X}=\boldsymbol{X}_i\right]\right|^2\leq\frac{1}{N}\sum_{i=1}^N\left(1-\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)}\right)^2\leq(1+\underline{c}^{-1})^2<\infty,\tag{S3.19}$$

$$\frac{1}{N}\sum_{i=1}^N\left\|\widehat{\mathbb{E}}\left[\frac{T}{\pi(\boldsymbol{Z};\gamma)^2}\nabla_\gamma\pi(\boldsymbol{Z};\gamma)\middle|\boldsymbol{X}=\boldsymbol{X}_i\right]\right\|^2\leq\frac{1}{N}\sum_{i=1}^N\left(\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)^2}\nabla_\gamma\pi(\boldsymbol{Z}_i;\gamma)\right)^2\leq\frac{C^2}{\underline{c}^4}<\infty,\tag{S3.20}$$

where $C := \sup_{\gamma\in\Gamma} \|\pi(\cdot;\gamma)\|_\infty < \infty$. Therefore, $\sup_{(\gamma,\theta)\in\Gamma\times\Theta} \|\nabla_{(\gamma,\theta)}\hat{Q}_N(\gamma,\theta)\| = O_p(1)$. Similarly, we have that $\sup_{(\gamma,\theta)\in\Gamma\times\Theta} \|\nabla_{(\gamma,\theta)}Q_0(\gamma,\theta)\| = O_p(1)$. All conditions imposed in Corollary 2.2 of Newey (1991) are satisfied, then it follows from Corollary 2.2 of Newey (1991) that

$$\sup_{(\gamma,\theta)\in\Gamma\times\Theta} |\hat{Q}_N(\gamma,\theta) - Q_0(\gamma,\theta)| \xrightarrow{p} 0,$$

which implies the consistency result $\|(\check{\gamma},\check{\theta}) - (\gamma_0,\theta_0)\| \xrightarrow{p} 0$.

Next, we establish the convergence rate $\|(\check{\gamma},\check{\theta}) - (\gamma_0,\theta_0)\| = O_p(N^{-1/2})$. Using the first order condition of optimization in Step I, we obtain that

$$\mathbf{G}_K(\check{\gamma},\check{\theta})^\top \cdot \widehat{\boldsymbol{W}}_0^{-1} \cdot \nabla_{\gamma,\theta}\mathbf{G}_K(\check{\gamma},\check{\theta}) = \mathbf{0}\ . \tag{S3.21}$$

An application of Mean Value Theorem yields:

$$\begin{aligned}\mathbf{0}_{1\times(p+1)} =& \frac{1}{\sqrt{N}}\boldsymbol{G}_K^\intercal(\gamma_0,\theta_0)\cdot\widehat{\boldsymbol{W}}_0^{-1}\cdot\left[\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\gamma_0,\theta_0)\right]\\ &+\left(\sqrt{N}(\check{\gamma}-\gamma_0)^\intercal, \sqrt{N}(\check{\theta}-\theta_0)\right)\cdot\left[\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\tilde{\gamma},\tilde{\theta})\right]^\intercal\cdot\widehat{\boldsymbol{W}}_0^{-1}\cdot\left[\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\tilde{\gamma},\tilde{\theta})\right]\\ &+\left[\frac{1}{N}\mathbf{G}_K(\tilde{\gamma},\tilde{\theta})\right]^\intercal\cdot\widehat{\boldsymbol{W}}_0^{-1}\cdot\left[\frac{1}{N}\nabla^2_{\gamma,\theta}\mathbf{G}_K(\tilde{\gamma},\tilde{\theta})\right]\left(\sqrt{N}(\check{\gamma}-\gamma_0)^\intercal, \sqrt{N}\cdot(\check{\theta}-\theta_0)\right)^\intercal\ ,\end{aligned} \tag{S3.22}$$

where $(\tilde{\gamma},\tilde{\theta})$ lies on the line joining $(\check{\gamma},\check{\theta})$ and (γ_0,θ_0). Obviously, in order to show $\|(\check{\gamma},\check{\theta}) - (\gamma_0,\theta_0)\| = O_p(N^{-1/2})$, it suffices to establish the following resutls:

$$\frac{1}{\sqrt{N}}\boldsymbol{G}_K^\intercal(\gamma_0,\theta_0)\cdot\widehat{\boldsymbol{W}}_0^{-1}\cdot\left[\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\gamma_0,\theta_0)\right] = O_p(1)\ , \tag{S3.23}$$

$$\left\{\left[\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\tilde{\gamma},\tilde{\theta})\right]^\intercal \cdot \widehat{\boldsymbol{W}}_0^{-1} \cdot \left[\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\tilde{\gamma},\tilde{\theta})\right]\right\}^{-1} = O_p(1)\ , \quad \text{(S3.24)}$$

$$\left[\frac{1}{N}\mathbf{G}_K(\tilde{\gamma},\tilde{\theta})\right]^\intercal \cdot \widehat{\boldsymbol{W}}_0^{-1} \cdot \left[\frac{1}{N}\nabla^2_{\gamma,\theta}\mathbf{G}_K(\tilde{\gamma},\tilde{\theta})\right] = o_p(1)\ . \quad \text{(S3.25)}$$

We first prove (S3.23). By computing the second moment and using Chebyshev's inequality, Inequality (S2.12), we can obtain that

$$\left\|\frac{1}{N}\boldsymbol{G}_K(\gamma_0,\theta_0)\right\| = O_p\left(\sqrt{\frac{K}{N}}\right)\ , \quad \text{(S3.26)}$$

$$\left\|\widehat{\boldsymbol{W}}_0 - \mathbb{E}[\widehat{\boldsymbol{W}}_0]\right\| = O_p\left(\sqrt{\frac{K^3}{N}}\right)\ , \quad \text{(S3.27)}$$

$$\left\|\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\gamma_0,\theta_0) - \boldsymbol{B}_{(K+1)\times(p+1)}\right\| = O_p\left(\sqrt{\frac{K}{N}}\right)\ , \quad \text{(S3.28)}$$

$$\left\|\frac{1}{N}\nabla^2_{\gamma,\theta}\boldsymbol{G}_K(\gamma_0,\theta_0) - \mathbb{E}\left[\frac{1}{N}\nabla^2_{\gamma,\theta}\boldsymbol{G}_K(\gamma_0,\theta_0)\right]\right\| = O_p\left(\sqrt{\frac{K}{N}}\right)\ . \quad \text{(S3.29)}$$

Using (S3.26) and (S3.27), we can deduce that

$$\frac{1}{\sqrt{N}}\boldsymbol{G}_K^\intercal(\gamma_0,\theta_0)\cdot\widehat{\boldsymbol{W}}_0^{-1}\cdot\left[\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\gamma_0,\theta_0)\right] = \frac{1}{\sqrt{N}}\boldsymbol{G}_K^\intercal(\gamma_0,\theta_0)\cdot\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\cdot\boldsymbol{B}_{(K+1)\times(p+1)} + o_p(1)\ .$$

Computing the variance of $N^{-1/2}\boldsymbol{G}_K^\intercal(\gamma_0,\theta_0)\cdot\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\cdot\boldsymbol{B}_{(K+1)\times(p+1)}$ yields:

$$\begin{aligned}
&\left\|Var\left(\frac{1}{\sqrt{N}}\boldsymbol{G}_K^\intercal(\gamma_0,\theta_0)\cdot\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\cdot\boldsymbol{B}_{(K+1)\times(p+1)}\right)\right\| \qquad \text{(S3.30)}\\
=&\left\|\boldsymbol{B}^\top_{(K+1)\times(p+1)}\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\mathbb{E}\left[g_K(T,\boldsymbol{Z};\gamma_0,\theta_0)^{\otimes 2}\right]\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\boldsymbol{B}_{(K+1)\times(p+1)}\right\|\\
=&\mathbb{E}\left[\left\|\boldsymbol{B}^\top_{(K+1)\times(p+1)}\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}g_K(T,\boldsymbol{Z};\gamma_0,\theta_0)\right\|^2\right]\\
=&\mathbb{E}\left[\left\|\mathbb{E}\left[\frac{\nabla_\gamma\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^\top\right]\mathbb{E}\left[u_K(\boldsymbol{X})^{\otimes 2}\right]^{-1}u_K(\boldsymbol{X})\left\{1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right\}\right.\right.\\
&\qquad\left.\left.+\,\mathbb{E}\left[\frac{\nabla_\gamma\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}\right]\left\{\theta_0-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right\}\right\|^2\right]\\
&+\mathbb{E}\left[\left|\theta_0-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right|^2\right]
\end{aligned}$$

$$
\begin{aligned}
&\leq 2 \cdot \mathbb{E}\left[\left\|\mathbb{E}\left[\frac{\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)} u_K(\boldsymbol{X})^\top\right] \mathbb{E}\left[u_K(\boldsymbol{X})^{\otimes 2}\right]^{-1} u_K(\boldsymbol{X})\right\|^2 \cdot \left\{1 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right\}^2\right] \\
&\quad + 2 \cdot \left\|\mathbb{E}\left[\frac{\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}\right]\right\|^2 \cdot \mathbb{E}\left[\left\{\theta_0 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)} U(\boldsymbol{Z})\right\}^2\right] \\
&\quad + \mathbb{E}\left[\left|\theta_0 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)} U(\boldsymbol{Z})\right|^2\right] \\
&\leq 2\left(1 + \underline{c}^{-1}\right)^2 \cdot \mathbb{E}\left[\left\|\mathbb{E}\left[\frac{\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)} u_K(\boldsymbol{X})^\top\right] \mathbb{E}\left[u_K(\boldsymbol{X})^{\otimes 2}\right]^{-1} u_K(\boldsymbol{X})\right\|^2\right] + O(1) + O(1).
\end{aligned}
$$

Note that $\mathbb{E}\left[\frac{\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)} u_K(\boldsymbol{X})^\top\right] \mathbb{E}\left[u_K(\boldsymbol{X})^{\otimes 2}\right]^{-1} u_K(\boldsymbol{X})$ is the L^2-projection of $\frac{\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}$ on the space spanned by $u_K(\boldsymbol{X})$, which implies that

$$
\mathbb{E}\left[\left\|\mathbb{E}\left[\frac{\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)} u_K(\boldsymbol{X})^\top\right] \mathbb{E}\left[u_K(\boldsymbol{X})^{\otimes 2}\right]^{-1} u_K(\boldsymbol{X})\right\|^2\right] \leq \mathbb{E}\left[\left\|\frac{\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}\right\|^2\right] < \infty.
$$

Then we have

$$
\left\|Var\left(\frac{1}{\sqrt{N}} \boldsymbol{G}_K^\intercal(\gamma_0,\theta_0) \cdot \mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1} \cdot \boldsymbol{B}_{(K+1)\times(p+1)}\right)\right\| = O(1),
$$

which implies (S3.23) by Chebyshev's inequality.

We next consider to prove (S3.24). Note that

$$
\left[\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\gamma,\theta)\right]^\intercal \cdot \widehat{\boldsymbol{W}}_0^{-1} \cdot \left[\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\gamma,\theta)\right] = \begin{pmatrix} \widehat{\boldsymbol{M}}^{(1)}_{p\times p}(\gamma), & \widehat{\boldsymbol{M}}^{(2)}_{p\times 1}(\gamma) \\ \left(\widehat{\boldsymbol{M}}^{(2)}_{p\times 1}(\gamma)\right)^\top, & 1 \end{pmatrix},
$$

where

$$
\begin{aligned}
\widehat{\boldsymbol{M}}^{(1)}_{p\times p}(\gamma) &:= \frac{1}{N}\sum_{i=1}^N \widehat{\mathbb{E}}\left[\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)^2}\nabla_\gamma \pi(\boldsymbol{Z}_i;\gamma)\middle|\boldsymbol{X}_i\right]^{\otimes 2} + \left\{\frac{1}{N}\sum_{i=1}^N \frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)^2} U(\boldsymbol{Z}_i)\nabla_\gamma \pi(\boldsymbol{Z}_i;\gamma)\right\}^{\otimes 2}, \\
\widehat{\boldsymbol{M}}^{(2)}_{p\times 1}(\gamma) &:= \frac{1}{N}\sum_{i=1}^N \frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)^2} U(\boldsymbol{Z}_i)\nabla_\gamma \pi(\boldsymbol{Z}_i;\gamma).
\end{aligned}
$$

By the least square projection property and Assumption 6, we have that for all $\gamma \in \Gamma$,

$$\begin{aligned}\left\|\widehat{\boldsymbol{M}}^{(1)}_{p\times p}(\gamma)\right\| \leq & \frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\mathbb{E}}\left[\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)^2}\nabla_\gamma \pi(\boldsymbol{Z}_i;\gamma)\middle|\boldsymbol{X}_i\right]\right\|^2 + \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)^2}U(\boldsymbol{Z}_i)\nabla_\gamma \pi(\boldsymbol{Z}_i;\gamma)\right\|^2 \\ \leq & \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)^2}\nabla_\gamma \pi(\boldsymbol{Z}_i;\gamma)\right\|^2 + \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)^2}U(\boldsymbol{Z}_i)\nabla_\gamma \pi(\boldsymbol{Z}_i;\gamma)\right\|^2 \\ \leq & \frac{C^2}{\underline{c}^4} + \frac{C^2}{\underline{c}^4}\cdot\frac{1}{N}\sum_{i=1}^{N}|U(\boldsymbol{Z}_i)|^2 = O_p(1),\end{aligned}$$

and

$$\|\widehat{\boldsymbol{M}}^{(2)}_{p\times 1}(\gamma)\| \leq \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma)^2}U(\boldsymbol{Z}_i)\nabla_\gamma \pi(\boldsymbol{Z}_i;\gamma)\right\|^2 \leq \frac{C^2}{\underline{c}^4}\cdot\frac{1}{N}\sum_{i=1}^{N}|U(\boldsymbol{Z}_i)|^2 = O_p(1).$$

Therefore, we can justify (S3.24).

Proving (S3.25) is similar to proving $\sup_{(\gamma,\theta)\in\Gamma\times\Theta}|\hat{Q}_N(\gamma,\theta)-Q_0(\gamma,\theta)| \xrightarrow{p} 0$. We can first show that for each fixed $(\gamma,\theta)\in\Gamma\times\Theta$, $[N^{-1}\mathbf{G}_K(\gamma,\theta)]^\intercal\cdot\widehat{\boldsymbol{W}}_0^{-1}\cdot\left[N^{-1}\nabla^2_{\gamma,\theta}\mathbf{G}_K(\gamma,\theta)\right]-\mathbb{E}\left[g_K(T,\boldsymbol{Z};\gamma,\theta)\right]^\intercal\cdot\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\cdot\mathbb{E}\left[\nabla^2_{\gamma,\theta}g_K(T,\boldsymbol{Z};\gamma,\theta)\right] = o_p(1)$, then strengthen to $\sup_{(\gamma,\theta)\in\Gamma\times\Theta}\ |[N^{-1}\mathbf{G}_K(\gamma,\theta)]^\intercal\cdot\widehat{\boldsymbol{W}}_0^{-1}\cdot[N^{-1}\nabla^2_{\gamma,\theta}\mathbf{G}_K(\gamma,\theta)]-\mathbb{E}[g_K(T,\boldsymbol{Z};\gamma,\theta)]^\intercal\cdot\mathbb{E}[\widehat{\boldsymbol{W}}_0]^{-1}\cdot\mathbb{E}[\nabla^2_{\gamma,\theta}g_K(T,\boldsymbol{Z};\gamma,\theta)]| = o_p(1)$. Then in light of the facts that $\|(\tilde{\gamma},\tilde{\theta})-(\gamma_0,\theta_0)\| \leq \|(\check{\gamma},\check{\theta})-(\gamma_0,\theta_0)\| \xrightarrow{p} 0$ and $\mathbb{E}[g_K(T,\boldsymbol{Z};\gamma_0,\theta_0)] = 0$, we can obtain (S3.25).

Finally, by combining (S3.23), (S3.24), (S3.25), and (S3.22), we can conclude our desired result $\|(\check{\gamma},\check{\theta})-(\gamma_0,\theta_0)\| = O_p(N^{-1/2})$.

S4 Proof of Theorem 2

The consistency result $\|(\bar{\gamma}, \bar{\theta}) - (\gamma_0, \theta_0)\| \xrightarrow{p} 0$ holds by using a similar argument of showing $\|(\check{\gamma}, \check{\theta}) - (\gamma_0, \theta_0)\| \xrightarrow{p} 0$ in Theorem 1.

We next show the asymptotic normality for the infeasible estimator $(\bar{\gamma}, \bar{\theta})$. Using the first order condition of optimization, we can obtain that

$$\mathbf{G}_K(\bar{\gamma}, \bar{\theta})^T \cdot \boldsymbol{D}^{-1}_{(K+1)\times(K+1)} \cdot \nabla_{\gamma,\theta}\mathbf{G}_K(\bar{\gamma}, \bar{\theta}) = \mathbf{0} \ ,$$

then an application of Mean Value Theorem yields:

$$\begin{aligned}\mathbf{0}_{1\times(p+1)} =& \frac{1}{\sqrt{N}}\boldsymbol{G}_K^{\intercal}(\gamma_0, \theta_0) \cdot \boldsymbol{D}^{-1}_{(K+1)\times(K+1)} \cdot \left[\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\gamma_0, \theta_0)\right] \\ &+ \left(\sqrt{N}(\bar{\gamma} - \gamma_0)^{\intercal}, \sqrt{N}(\bar{\theta} - \theta_0)\right) \cdot \left[\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\gamma^*, \theta^*)\right]^{\intercal} \cdot \boldsymbol{D}^{-1}_{(K+1)\times(K+1)} \cdot \left[\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\gamma^*, \theta^*)\right] \\ &+ \left[\frac{1}{N}\mathbf{G}_K(\gamma^*, \theta^*)\right]^{\intercal} \cdot \boldsymbol{D}^{-1}_{(K+1)\times(K+1)} \cdot \left[\frac{1}{N}\nabla^2_{\gamma,\theta}\mathbf{G}_K(\gamma^*, \theta^*)\right] \left(\sqrt{N}(\bar{\gamma} - \gamma_0)^{\intercal}, \sqrt{N} \cdot (\bar{\theta} - \theta_0)\right)^{\intercal} \ .\end{aligned} \tag{S4.31}$$

where (γ^*, θ^*) lies on the line joining $(\bar{\gamma}, \bar{\theta})$ and (γ_0, θ_0). Note that the expression (S4.31) has the same structure as (S3.22), except for that the weighting matrix used in (S4.31) is $\boldsymbol{D}^{-1}_{(K+1)\times(K+1)}$ while the weighting matrix used in (S3.22) is $\widehat{\boldsymbol{W}}_0^{-1}$. Using a similar argument of showing $\sup_{(\gamma,\theta)\in\Gamma\times\Theta} |\hat{Q}_N(\gamma, \theta) - Q_0(\gamma, \theta)| \xrightarrow{p} 0$ and (S3.25), we can obtain the following results:

$$\begin{aligned}&\left[\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\gamma^*, \theta^*)\right]^{\intercal} \cdot \boldsymbol{D}^{-1}_{(K+1)\times(K+1)} \cdot \left[\frac{1}{N}\nabla_{\gamma,\theta}\boldsymbol{G}_K(\gamma^*, \theta^*)\right] \\ =&(\boldsymbol{B}_{(K+1)\times(p+1)})^{\intercal} \cdot \boldsymbol{D}^{-1}_{(K+1)\times(K+1)} \cdot \boldsymbol{B}_{(K+1)\times(p+1)} + o_p(1)\end{aligned} \tag{S4.32}$$

and

$$\left[\frac{1}{N}\mathbf{G}_K(\gamma^*, \theta^*)\right]^{\intercal} \cdot \boldsymbol{D}^{-1}_{(K+1)\times(K+1)} \cdot \left[\frac{1}{N}\nabla^2_{\gamma,\theta}\mathbf{G}_K(\gamma^*, \theta^*)\right] = o_p(1) \ . \tag{S4.33}$$

Combining (S3.28), (S4.31), (S4.32), and (S4.33) together, we can deduce that

$$\begin{pmatrix} \sqrt{N}(\bar{\gamma}-\gamma_0) \\ \sqrt{N}(\bar{\theta}-\theta_0) \end{pmatrix} = -\left\{(\boldsymbol{B}_{(K+1)\times(p+1)})^{\intercal}\cdot \boldsymbol{D}^{-1}_{(K+1)\times(K+1)}\cdot \boldsymbol{B}_{(K+1)\times(p+1)}\right\}^{-1} \cdot (\boldsymbol{B}_{(K+1)\times(p+1)})^{\intercal}\boldsymbol{D}^{-1}_{(K+1)\times(K+1)}\cdot \frac{1}{\sqrt{N}}\boldsymbol{G}_K(\gamma_0,\theta_0)+o_p(1)\ .$$

Then

$$Cov\left(\begin{pmatrix} \sqrt{N}(\bar{\gamma}-\gamma_0) \\ \sqrt{N}(\bar{\theta}-\theta_0) \end{pmatrix}\right) = \left((\boldsymbol{B}_{(K+1)\times(p+1)})^{\intercal}\cdot \boldsymbol{D}^{-1}_{(K+1)\times(K+1)}\cdot \boldsymbol{B}_{(K+1)\times(p+1)}\right)^{-1}+o(1)$$
$$=\boldsymbol{V}_K+o(1)\ .$$

Therefore, we can obtain that

$$\sqrt{N}\boldsymbol{V}_K^{-1/2}\begin{pmatrix} \bar{\gamma}-\gamma_0 \\ \bar{\theta}-\theta_0 \end{pmatrix} = -\left\{\boldsymbol{B}^{\intercal}_{(K+1)\times(p+1)}\cdot \boldsymbol{D}^{-1}_{(K+1)\times(K+1)}\cdot \boldsymbol{B}_{(K+1)\times(p+1)}\right\}^{-1/2} \cdot \boldsymbol{B}^{\intercal}_{(K+1)\times(p+1)}\boldsymbol{D}^{-1}_{(K+1)\times(K+1)}\cdot \frac{1}{\sqrt{N}}\boldsymbol{G}_K(\gamma_0,\theta_0)+o_p(1). \tag{S4.34}$$

We next show that the normalized estimator $\sqrt{N}\boldsymbol{V}_K^{-1/2}\begin{pmatrix} \bar{\gamma}-\gamma_0 \\ \bar{\theta}-\theta_0 \end{pmatrix}$ converges to the normal distribution. The key part of the proof is to verify the Lindeberg type conditions imposed in Eicker (1966). Note that

$\frac{1}{\sqrt{N}}\boldsymbol{G}_K(\gamma_0,\theta_0)$ can be written as:

$$\frac{1}{\sqrt{N}}\boldsymbol{G}_K(\gamma_0,\theta_0) = \begin{pmatrix} \frac{1}{\sqrt{N}}\sum_{i=1}^N\left[1-\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma_0)}\right]u_K(\boldsymbol{X}_i) \\ \frac{1}{\sqrt{N}}\sum_{i=1}^N\left[\theta_0-\frac{T_i}{\pi(\boldsymbol{Z}_i;\gamma_0)}U(\boldsymbol{Z}_i)\right] \end{pmatrix}$$
$$=\boldsymbol{A}_{(K+1)\times N(K+1)}\cdot\mathcal{E}_{N(K+1)\times 1}\ ,$$

where

$$\boldsymbol{A}_{(K+1)\times N(K+1)} := \begin{pmatrix} \frac{1}{\sqrt{N}}\boldsymbol{1}_{1\times N}, & \boldsymbol{0}_{1\times N}, & \boldsymbol{0}_{1\times N}, & \cdots\boldsymbol{0}_{1\times N}, & \boldsymbol{0}_{1\times N} \\ \boldsymbol{0}_{1\times N}, & \frac{1}{\sqrt{N}}\boldsymbol{1}_{1\times N}, & \boldsymbol{0}_{1\times N}, & \cdots\boldsymbol{0}_{1\times N}, & \boldsymbol{0}_{1\times N} \\ \vdots & \vdots & \vdots & \vdots & \\ \boldsymbol{0}_{1\times N}, & \boldsymbol{0}_{1\times N} & \boldsymbol{0}_{1\times N}, & \frac{1}{\sqrt{N}}\boldsymbol{1}_{1\times N}, & \boldsymbol{0}_{1\times N} \\ \boldsymbol{0}_{1\times N}, & \boldsymbol{0}_{1\times N} & \boldsymbol{0}_{1\times N}, & \cdots\boldsymbol{0}_{1\times N} & \frac{1}{\sqrt{N}}\boldsymbol{1}_{1\times N} \end{pmatrix}$$

is a $(K+1)\times N(K+1)$ matrix, and

$$\mathcal{E}_{N(K+1)\times 1} := (\boldsymbol{v}_{1\times N}(1),\ldots,\boldsymbol{v}_{1\times N}(K),\boldsymbol{w}_{1\times N})^\top\ ,$$
$$\boldsymbol{v}_{1\times N}(k) := \left(\left[1-\frac{T_1}{\pi(\boldsymbol{Z}_1;\gamma_0)}\right]u_{kK}(\boldsymbol{X}_1),\ldots,\left[1-\frac{T_N}{\pi(\boldsymbol{Z}_N;\gamma_0)}\right]u_{kK}(\boldsymbol{X}_N)\right),$$
$$\boldsymbol{w}_{1\times N} := \left(\theta_0-\frac{T_1}{\pi(\boldsymbol{Z}_1;\gamma_0)}U(\boldsymbol{Z}_1),\cdots,\theta_0-\frac{T_N}{\pi(\boldsymbol{Z}_N;\gamma_0)}U(\boldsymbol{Z}_N)\right)$$

for $k\in\{1,\ldots,K\}$, and $\boldsymbol{1}_{1\times N}$ (resp. $\boldsymbol{0}_{1\times N}$) denotes a N-dimensional row vector whose elements are all of 1's (resp. 0's). From Eicker (1966), the following Lindeberg type conditions are sufficient to ensure $\sqrt{N}\boldsymbol{V}_K^{-1/2}(\bar{\gamma}-\gamma_0, \bar{\theta}-\theta_0)^\top \xrightarrow{d} N(0,I_{(p+1)\times(p+1)})$, namely,

1. $\max_{i\in\{1,...,N(K+1)\}} \boldsymbol{a}_i^\top \left(\boldsymbol{A}_{(K+1)\times N(K+1)}\boldsymbol{A}_{(K+1)\times N(K+1)}^\top\right)^{-1}\boldsymbol{a}_i \to 0$, where $\boldsymbol{a}_i$ is the i^{th} column of $\boldsymbol{A}_{(K+1)\times N(K+1)}$;

2. both $\sup_{k\in\{1,...,K\}} \mathbb{E}\left[\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)^2 u_{kK}(\boldsymbol{X})^2 I\left(\left|\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)u_{kK}(\boldsymbol{X})\right|\geq s\right)\right]\to 0$ and

$$\mathbb{E}\left[\left(\theta_0-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2 I\left(\left|\theta_0-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right|\geq s\right)\right]\to 0$$

as $s\to\infty$;

3. both $\inf_{k\in\{1,...,K\}} \mathbb{E}\left[\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)^2 u_{kK}(\boldsymbol{X})^2\right]>0$ and $\mathbb{E}\left[\left(\theta_0-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right]>0$.

Conditions 1 is natually satisfied by the definition of $\boldsymbol{A}_{(K+1)\times N(K+1)}$. Condition 2 holds because

$$\begin{aligned}&\mathbb{E}\left[\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)^2 u_{kK}(\boldsymbol{X})^2 I\left(\left|\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)u_{kK}(\boldsymbol{X})\right|\geq s\right)\right]\\ \leq&\left(1+\frac{1}{\underline{c}}\right)^2\cdot\mathbb{E}\left[u_{kK}(\boldsymbol{X})^2 I\left(|u_{kK}(\boldsymbol{X})|\geq s\left(1+\frac{1}{\underline{c}}\right)^{-1}\right)\right]\xrightarrow{s\to\infty}0\ ,\end{aligned}$$

where $\underline{c}>0$ is lower bound of the propensity socre $\pi(\boldsymbol{z};\gamma_0)$ (Assumption 6), and the last convergence holds from the fact $\sup_{k\in\{1,..,K\}}\mathbb{E}[u_{kK}^2(\boldsymbol{X})]\leq\overline{a}$ (see (S1.9)) and Dominated Convergence Theorem; the second part in Condition 2 also follows from Assumption 6 and Dominated Convergence Theorem. Condition 3 holds because

$$\mathbb{E}\left[\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)^2 u_K(\boldsymbol{X})^{\otimes 2}\right]=\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z})}{\pi(\boldsymbol{Z})}u_K(\boldsymbol{X})^{\otimes 2}\right]\geq\frac{1-\bar{c}}{\bar{c}}\mathbb{E}\left[u_K(\boldsymbol{X})^{\otimes 2}\right]\geq\frac{(1-\bar{c})\underline{a}}{\bar{c}}\cdot I_{K\times K}\ ,$$

which implies that $\inf_{k\in\{1,\dots,K\}} \mathbb{E}\left[\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)^2 u_{kK}(\boldsymbol{X})^2\right] > 0$; the second part in Condition 3 is obvious. Therefore, Conditions 1, 2, and 3 are all satisfied, then we can conclude our desired result that

$$\sqrt{N}\boldsymbol{V}_K^{-1/2}\left(\begin{array}{c}\bar{\gamma}-\gamma_0, \bar{\theta}-\theta_0\end{array}\right)^\top \xrightarrow{d} N\left(0, I_{(p+1)\times(p+1)}\right).$$

S5 Proof of Theorem 3

Morikawa and Kim (2016) show that the efficient variance bounds of (γ_0, θ_0) is $\boldsymbol{V}_{eff} := E\left[\boldsymbol{S}_{eff}(T, \mathbf{Z}; \gamma_0, \theta_0)^{\otimes 2}\right]^{-1}$, where $\boldsymbol{S}_{eff} = (\boldsymbol{S}_1^\top, S_2)^\top$ and $\boldsymbol{S}_1$, S_2 are defined in (S1.3) and (S1.4) respectively. Let $\boldsymbol{V}_{\gamma_0}$ (resp. V_{θ_0}) be the efficient variance bound of γ_0 (resp. θ_0). After some simple computation, we can find out

$$\boldsymbol{V}_{\gamma_0} = E\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)} m(\boldsymbol{X})^{\otimes 2}\right]^{-1}$$

and

$$V_{\theta_0} = Var\left(S_2(T, \boldsymbol{Z}; \gamma_0, \theta_0) - \kappa^\intercal \boldsymbol{S}_1(T, \boldsymbol{Z}; \gamma_0)\right) .$$

where

$$\kappa^\intercal = E\left[\frac{\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)^\intercal}{\pi(\boldsymbol{Z};\gamma_0)}\{R(\boldsymbol{Z}) - U(\boldsymbol{X})\}\right] \cdot E\left[\frac{m(\boldsymbol{X})}{\pi(\boldsymbol{Z};\gamma_0)} \nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)^\intercal\right]^{-1} .$$

From Theorem 2 we know have that $\sqrt{N}\boldsymbol{V}_K^{-1/2}\left(\begin{array}{c}\bar{\gamma}-\gamma_0, \bar{\theta}-\theta_0\end{array}\right)^\top \xrightarrow{d} N\left(0, I_{(p+1)\times(p+1)}\right)$, therefore to prove Theorem 3, it suffices to show $\boldsymbol{V}_K$

converges to the efficient variance bound of (γ_0, θ_0).

Since $\boldsymbol{V}_K = \left\{\boldsymbol{B}^{\intercal}_{(K+1)\times(p+1)} \cdot \boldsymbol{D}^{-1}_{(K+1)\times(K+1)} \cdot \boldsymbol{B}_{(K+1)\times(p+1)}\right\}^{-1}$, we first find the expression of $\boldsymbol{V}_K$ without the inverse. Using the inverse matrix formula (S2.11), we can have:

$$\mathbf{D}^{-1}_{(K+1)\times(K+1)} = \begin{pmatrix} \boldsymbol{A}^{-1}_{K\times K} + \frac{1}{c}\boldsymbol{A}^{-1}_{K\times K}\boldsymbol{b}_K\boldsymbol{b}_K^{\intercal}\boldsymbol{A}^{-1}_{K\times K}, & -\frac{1}{c}\boldsymbol{A}^{-1}_{K\times K}\boldsymbol{b}_K \\ \\ -\frac{1}{c}\mathbf{b}_K^T\boldsymbol{A}^{-1}_{K\times K}, & \frac{1}{c} \end{pmatrix}, \tag{S5.35}$$

where

$$\boldsymbol{A}_{K\times K} := \mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^{\otimes 2}\right], \quad \boldsymbol{b}_K := \mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})u_K(\boldsymbol{X})\right], \tag{S5.36}$$

$$c := \mathbb{E}\left[\left(\theta_0 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right] - \boldsymbol{b}_K^T\boldsymbol{A}^{-1}_{K\times K}\boldsymbol{b}_K\ . \tag{S5.37}$$

Then we have

$$\boldsymbol{B}^{\intercal}_{(K+1)\times(p+1)} \cdot \boldsymbol{D}^{-1}_{(K+1)\times(K+1)} \cdot \boldsymbol{B}_{(K+1)\times(p+1)} = \begin{pmatrix} \tilde{\boldsymbol{A}}_{p\times p}, & \tilde{\boldsymbol{b}}_p \\ \\ \tilde{\boldsymbol{b}}_p^{\intercal} & \frac{1}{c} \end{pmatrix},$$

where

$$\tilde{\boldsymbol{A}}_{p\times p} := \mathbb{E}\left[\frac{\nabla_\gamma\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^{\intercal}\right] \cdot \boldsymbol{A}^{-1}_{K\times K} \cdot \mathbb{E}\left[u_K(\boldsymbol{X})\frac{\nabla_\gamma\pi(\boldsymbol{Z};\gamma_0)^{\intercal}}{\pi(\boldsymbol{Z};\gamma_0)}\right] + c\cdot\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal}, \tag{S5.38}$$

$$\tilde{\boldsymbol{b}}_p := -\frac{1}{c}\mathbb{E}\left[\frac{\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)} u_K(\boldsymbol{X})^\intercal\right] \cdot \boldsymbol{A}_{K\times K}^{-1} \cdot \boldsymbol{b}_K + \frac{1}{c}\mathbb{E}\left[\frac{\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)} U(\boldsymbol{Z})\right] . \tag{S5.39}$$

Using the matrix inversion formula (S2.11) again, we can obtain that

$$\begin{aligned}
\boldsymbol{V}_K &= \left(\boldsymbol{B}_{(K+1)\times(p+1)}^\intercal \cdot \boldsymbol{D}_{(K+1)\times(K+1)}^{-1} \cdot \boldsymbol{B}_{(K+1)\times(p+1)}\right)^{-1} \\
&= \begin{pmatrix} \tilde{\boldsymbol{A}}_{p\times p}^{-1} + \frac{1}{\tilde{c}}\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^\intercal\tilde{\boldsymbol{A}}_{p\times p}^{-1}, & -\frac{1}{\tilde{c}}\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p \\ \\ -\frac{1}{\tilde{c}}\tilde{\boldsymbol{b}}_p^\intercal\tilde{\boldsymbol{A}}_{p\times p}^{-1} & \frac{1}{\tilde{c}} \end{pmatrix},
\end{aligned}$$

where

$$\tilde{c} := \frac{1}{c} - \tilde{\boldsymbol{b}}_p^\intercal\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p .$$

Then we can obtain that

$$Var(\sqrt{N}(\bar{\gamma}-\gamma_0)) = \tilde{\boldsymbol{A}}_{p\times p}^{-1} + \frac{1}{\tilde{c}}\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^\intercal\tilde{\boldsymbol{A}}_{p\times p}^{-1} + o(1), \tag{S5.40}$$

$$Var(\sqrt{N}(\bar{\theta}-\theta_0)) = \frac{1}{\tilde{c}} + o(1) . \tag{S5.41}$$

In order to establish our Theorem 3, it suffices to show that (S5.40) and (S5.41) converge to $\boldsymbol{V}_{\gamma_0}$ and V_{θ_0} respectively. The proof is constituted of two parts:

Proof of Part (I)

We first show that (S5.40) **converges to** $\boldsymbol{V}_{\gamma_0}$ **as** N **goes to infinity**.

Note that

$$
\begin{aligned}
Var(\sqrt{N}(\bar{\gamma}-\gamma_0)) &= \tilde{\boldsymbol{A}}_{p\times p}^{-1} + \frac{1}{\tilde{c}}\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal}\tilde{\boldsymbol{A}}_{p\times p}^{-1} + o(1) \\
&= \left[\tilde{\boldsymbol{A}}_{p\times p} - c\cdot\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal}\right]^{-1} + o(1) \\
&= \left\{\mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^T\right]\cdot\boldsymbol{A}_{K\times K}^{-1}\cdot\mathbb{E}\left[u_K(\boldsymbol{X})\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)^T}{\pi(\boldsymbol{Z};\gamma_0)}\right]\right\}^{-1} + o(1) \quad \text{(using (S5.38))}\ ,
\end{aligned}
\tag{S5.42}
$$

where the second equality can be straightforwardly verified as follows:

$$
\begin{aligned}
&\left(\tilde{\boldsymbol{A}}_{p\times p}^{-1} + \frac{1}{\tilde{c}}\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal}\tilde{\boldsymbol{A}}_{p\times p}^{-1}\right)\cdot\left(\tilde{\boldsymbol{A}}_{p\times p} - c\cdot\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal}\right) \\
=&I_{p\times p} - c\cdot\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal} + \frac{1}{\tilde{c}}\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal} - \frac{c}{\tilde{c}}\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal}\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal} \\
=&I_{p\times p} - \frac{1}{\tilde{c}}\cdot\left[\tilde{c}c\cdot\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal} + c\cdot\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal}\cdot\tilde{\boldsymbol{b}}_p^{\intercal}\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\right] + \frac{1}{\tilde{c}}\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal} \\
=&I_{p\times p} - \frac{1}{\tilde{c}}\cdot\left[\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal}\right] + \frac{1}{\tilde{c}}\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal} \qquad \left(\text{since } c\tilde{c} = 1 - c\cdot\tilde{\boldsymbol{b}}_p^{\intercal}\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\right) \\
=&I_{p\times p}\ .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
Var(\sqrt{N}(\bar{\gamma}-\gamma_0)) =& \left\{\mathbb{E}\left[\frac{\nabla_\gamma\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^T\right]\cdot\boldsymbol{A}_{K\times K}^{-1}\cdot\mathbb{E}\left[u_K(\boldsymbol{X})\frac{\nabla_\gamma\pi(\boldsymbol{Z};\gamma_0)^T}{\pi(\boldsymbol{Z};\gamma_0)}\right]\right\}^{-1} \\
=&\mathbb{E}\left[f_K(\boldsymbol{Z})f_K(\boldsymbol{Z})^{\intercal}\right]^{-1} + o(1)\ ,
\end{aligned}
\tag{S5.43}
$$

where

$$
f_K(\boldsymbol{Z}) := -\,\mathbb{E}\left[\frac{\nabla_\gamma\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^{\intercal}\right]\cdot\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^{\otimes 2}\right]^{-1}u_K(\boldsymbol{X})\cdot\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)\ .
$$

Recalling the definitions of $\boldsymbol{S}_0(\boldsymbol{Z};\gamma_0)$ and $O(\boldsymbol{Z})$ in (S1.1), we can obtain that

$$
\begin{aligned}
f_K(\boldsymbol{Z}) =&\mathbb{E}\left[\boldsymbol{S}_0(\boldsymbol{Z};\gamma_0)O(\boldsymbol{Z})u_K(\boldsymbol{X})^{\intercal}\right]\cdot\mathbb{E}\left[O(\boldsymbol{Z})u_K(\boldsymbol{X})^{\otimes 2}\right]^{-1}u_K(\boldsymbol{X})\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right) \\
=&\left\{\mathbb{E}\left[\frac{\mathbb{E}[\boldsymbol{S}_0(\boldsymbol{Z};\gamma_0)O(\boldsymbol{Z})|\boldsymbol{X}]}{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}}\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}u_K(\boldsymbol{X})^{\intercal}\right]\cdot\mathbb{E}\left[\left(\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}u_K(\boldsymbol{X})\right)^{\otimes 2}\right]^{-1}\right.
\end{aligned}
$$

$$\cdot \sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]} u_K(\boldsymbol{X}) \Big\} \frac{1}{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}} \left(1 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)$$
$$= h_K(\boldsymbol{X}) \frac{1}{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}} \left(1 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right) ,$$

where

$$h_K(\boldsymbol{X}) := \mathbb{E}\left[\frac{\mathbb{E}[\boldsymbol{S}_0(\boldsymbol{Z};\gamma_0)O(\boldsymbol{Z})|\boldsymbol{X}]}{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}} \sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]} u_K(\boldsymbol{X})^\intercal\right] \cdot \mathbb{E}\left[\left(\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]} u_K(\boldsymbol{X})\right)^{\otimes 2}\right]^{-1} \cdot \sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]} u_K(\boldsymbol{X}) .$$

Note that $h_K(\boldsymbol{X})$ is the least square projection of $\frac{\mathbb{E}[\boldsymbol{S}_0(\boldsymbol{Z};\gamma_0)O(\boldsymbol{Z})|\boldsymbol{X}]}{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}}$ on the space linearly spanned by $\{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]} u_K(\boldsymbol{X})\}$, by Assumption 3 and Lemma 1, we can have

$$\mathbb{E}\left[\left\| h_K(\boldsymbol{X}) - \frac{\mathbb{E}[\boldsymbol{S}_0(\boldsymbol{Z};\gamma_0)O(\boldsymbol{Z})|\boldsymbol{X}]}{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}} \right\|^2\right] = O\left(K^{-\frac{2s}{r}}\right) = o(1) . \quad \text{(S5.44)}$$

Therefore, we can have that

$$\begin{aligned}
\mathbb{E}\left[f_K(\boldsymbol{Z}) f_K(\boldsymbol{Z})^\top\right] &= \mathbb{E}\left[h_K(\boldsymbol{X}) h_K(\boldsymbol{X})^\top \frac{1}{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]} \left(1 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)^2\right] \\
&= \mathbb{E}\left[h_K(\boldsymbol{X}) h_K(\boldsymbol{X})^\top\right] \qquad \left(\text{since } \mathbb{E}\left[\left(1 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)^2 \Big| \boldsymbol{X}\right] = \mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]\right) \\
&\to \mathbb{E}\left[\frac{\mathbb{E}[\boldsymbol{S}_0(\boldsymbol{Z};\gamma_0)O(\boldsymbol{Z})|\boldsymbol{X}] \cdot \mathbb{E}[\boldsymbol{S}_0(\boldsymbol{Z};\gamma_0)O(\boldsymbol{Z})|\boldsymbol{X}]^\top}{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}\right] \\
&= \mathbb{E}\left[m(\boldsymbol{X}) m(\boldsymbol{X})^\top \left(1 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)^2\right] .
\end{aligned}$$

Then in light of (S5.43) we can obtain our desired result:

$$Var(\sqrt{N}(\bar{\gamma} - \gamma_0)) \to \mathbb{E}\left[\left(1 - \frac{T}{\pi(\mathbf{Z};\gamma_0)}\right)^2 \cdot m(\boldsymbol{X})^{\otimes 2}\right]^{-1} = \boldsymbol{V}_{\gamma_0} .$$

Proof of Part (II)

Next, we show that (S5.41) **converges to** V_{θ_0} **as** N **goes to infinity.**

Applying the matrix inverse formula (S2.10) to (S5.41), we have

$$
\begin{aligned}
&Var(\sqrt{N}(\bar{\theta}-\theta_0)) = \left[\frac{1}{c} - \tilde{\boldsymbol{b}}_p^{\intercal}\tilde{\boldsymbol{A}}_{p\times p}^{-1}\tilde{\boldsymbol{b}}_p\right]^{-1} + o(1)\\
=&c - c\cdot\tilde{\boldsymbol{b}}_p^{\intercal}\cdot\left(-\tilde{\boldsymbol{A}}_{p\times p} + c\cdot\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal}\right)^{-1}\tilde{\boldsymbol{b}}_p\cdot c + o(1) \qquad \text{(by (S2.10))}\\
=&c + \left(\mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^{\intercal}\right]\cdot\boldsymbol{A}_{K\times K}^{-1}\cdot\boldsymbol{b}_K - \mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right]\right)^{\intercal} + o(1)\\
&\cdot\left\{\mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^{\intercal}\right]\cdot\boldsymbol{A}_{K\times K}^{-1}\cdot\mathbb{E}\left[u_K(\boldsymbol{X})\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)^{\intercal}}{\pi(\boldsymbol{Z};\gamma_0)}\right]\right\}^{-1} + o(1)\\
&\cdot\left(\mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^{\intercal}\right]\cdot\boldsymbol{A}_{K\times K}^{-1}\cdot\boldsymbol{b}_K - \mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right]\right) + o(1)\ ,
\end{aligned}
\tag{S5.45}
$$

where the last equality follows from the definitions of $\tilde{\boldsymbol{A}}_{p\times p}$ and $\tilde{\boldsymbol{b}}_p$ in (S5.38) and (S5.39), namely,

$$
\begin{aligned}
&\tilde{\boldsymbol{A}}_{p\times p} - c\cdot\tilde{\boldsymbol{b}}_p\tilde{\boldsymbol{b}}_p^{\intercal} = \mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^T\right]\cdot\boldsymbol{A}_{K\times K}^{-1}\cdot\mathbb{E}\left[u_K(\boldsymbol{X})\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)^T}{\pi(\boldsymbol{Z};\gamma_0)}\right]\ ,\\
&\tilde{\boldsymbol{b}}_p\cdot c = -\mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^T\right]\cdot\boldsymbol{A}_{K\times K}^{-1}\cdot\boldsymbol{b}_K + \mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right]\ .
\end{aligned}
$$

In the following we show that

$$
c = \mathbb{E}\left[\left(\theta_0 - \frac{T}{\pi(\boldsymbol{Z})}U(\boldsymbol{Z})\right)^2\right] - \mathbb{E}\left[\left\{\left(1 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)R(\boldsymbol{X})\right\}^2\right] + o(1)\ ; \tag{S5.46}
$$

$$
\mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^T\right]\cdot\boldsymbol{A}_{K\times K}^{-1}\cdot\boldsymbol{b}_K - \mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right] = \mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}(R(\boldsymbol{X}) - U(\boldsymbol{Z}))\right] + o(1)\ ; \tag{S5.47}
$$

$$
\left\{\mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^T\right]\cdot\boldsymbol{A}_{K\times K}^{-1}\cdot\mathbb{E}\left[u_K(\boldsymbol{X})\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)^T}{\pi(\boldsymbol{Z};\gamma_0)}\right]\right\}^{-1} = V_{\gamma_0} + o(1)\ . \tag{S5.48}
$$

where c is defined in (S5.37).

For the term (S5.46): Note that

$$
\begin{aligned}
c =&\mathbb{E}\left[\left(\theta_0 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right] \\
&- \mathbb{E}\left[O(\mathbf{Z})U(\boldsymbol{Z})u_K(\boldsymbol{X})^\intercal\right]\cdot\mathbb{E}\left[O(\mathbf{Z})u_K(\boldsymbol{X})^{\otimes 2}\right]^{-1}\cdot\mathbb{E}\left[O(\mathbf{Z})U(\boldsymbol{Z})u_K(\boldsymbol{X})\right] \quad \text{(by definition (S5.37))} \\
=&\mathbb{E}\left[\left(\theta_0 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right] \\
&- \mathbb{E}\left[\left\{\mathbb{E}\left[O(\mathbf{Z})U(\boldsymbol{Z})u_K(\boldsymbol{X})^\intercal\right]\cdot\mathbb{E}\left[O(\mathbf{Z})u_K(\boldsymbol{X})^{\otimes 2}\right]^{-1}\cdot u_K(\boldsymbol{X})\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)\right\}^2\right] \\
=&\mathbb{E}\left[\left(\theta_0 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right] \\
&- \mathbb{E}\left[\left\{\mathbb{E}\left[\mathbb{E}\left[O(\boldsymbol{Z})U(\boldsymbol{Z})|\boldsymbol{X}\right]\cdot u_K(\boldsymbol{X})^\intercal\right]\cdot\mathbb{E}\left[\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]\cdot u_K(\boldsymbol{X})^{\otimes 2}\right]^{-1}\cdot u_K(\boldsymbol{X})\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)\right\}^2\right] \\
=&\mathbb{E}\left[\left(\theta_0 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right] \\
&- \mathbb{E}\left[\left\{\mathbb{E}\left[\frac{\mathbb{E}\left[O(\boldsymbol{Z})U(\boldsymbol{Z})|\boldsymbol{X}\right]}{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}}\cdot\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}u_K(\boldsymbol{X})^T\right]\cdot\mathbb{E}\left[\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]\cdot u_K(\boldsymbol{X})^{\otimes 2}\right]^{-1}\right.\right. \\
&\qquad\left.\left.\times\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}u_K(\boldsymbol{X})\cdot\frac{1}{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}}\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)\right\}^2\right].
\end{aligned}
$$

Considering the last term in above expression, since

$$
\mathbb{E}\left[\frac{\mathbb{E}\left[O(\boldsymbol{Z})U(\boldsymbol{Z})|\boldsymbol{X}\right]}{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}}\cdot\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}u_K(\boldsymbol{X})^\intercal\right]\cdot\mathbb{E}\left[\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]\cdot u_K(\boldsymbol{X})^{\otimes 2}\right]^{-1}\cdot\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}u_K(\boldsymbol{X})
$$

is the L^2-projection of $\frac{\mathbb{E}[O(\boldsymbol{Z})U(\boldsymbol{Z})|\boldsymbol{X}]}{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}}$ on the space linearly spanned by $\{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}u_K(\boldsymbol{X})\}$, by Assumption 3 and Lemma 1, we can have

$$
\begin{aligned}
\mathbb{E}\Bigg[\Bigg|&\mathbb{E}\left[\frac{\mathbb{E}\left[O(\boldsymbol{Z})U(\boldsymbol{Z})|\boldsymbol{X}\right]}{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}}\cdot\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}u_K(\boldsymbol{X})^\intercal\right]\cdot\mathbb{E}\left[\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]\cdot u_K(\boldsymbol{X})^{\otimes 2}\right]^{-1}\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}u_K(\boldsymbol{X}) \\
&- \frac{\mathbb{E}\left[O(\boldsymbol{Z})U(\boldsymbol{Z})|\boldsymbol{X}\right]}{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}}\Bigg|^2\Bigg] = O(K^{-\frac{2s}{r}}) = o(1)\,. \qquad \text{(S5.49)}
\end{aligned}
$$

Then we can have that

$$
c \to \mathbb{E}\left[\left(\theta_0 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right] - \mathbb{E}\left[\left\{\frac{\mathbb{E}\left[O(\boldsymbol{Z})U(\boldsymbol{Z})|\boldsymbol{X}\right]}{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}}\cdot\frac{1}{\sqrt{\mathbb{E}[O(\boldsymbol{Z})|\boldsymbol{X}]}}\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)\right\}^2\right]
$$

$$=\mathbb{E}\left[\left(\theta_0 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right] - \mathbb{E}\left[\left\{\left(1 - \frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)R(\boldsymbol{X})\right\}^2\right] ,$$

which is (S5.46).

For the term (S5.47). Consider the term

$$\mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^\intercal\right]\cdot \boldsymbol{A}_{K\times K}^{-1}\cdot \boldsymbol{b}_K - \mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right] .$$

Using (S5.49), we can have

$$\begin{aligned}
&\mathbb{E}\left[\frac{\nabla_\gamma\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^\top\right]\cdot \boldsymbol{A}_{K\times K}^{-1}\cdot \boldsymbol{b}_K\\
=&\mathbb{E}\left[O(\boldsymbol{Z})U(\boldsymbol{Z})u_K(\boldsymbol{X})^\top\right]\cdot\mathbb{E}\left[O(\boldsymbol{Z})u_K(\boldsymbol{X})^{\otimes 2}\right]^{-1}\cdot\mathbb{E}\left[u_K(\boldsymbol{X})\frac{\nabla_\gamma\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}\right]\\
=&\mathbb{E}\left[R(\boldsymbol{X})\frac{\nabla_\gamma\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}\right]+o(1) \qquad \text{(using (S5.49))} .
\end{aligned}$$

Therefore, we can obtain that

$$\mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^\intercal\right]\cdot \boldsymbol{A}_{K\times K}^{-1}\cdot \boldsymbol{b}_K - \mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right] \to \mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}(R(\boldsymbol{X})-U(\boldsymbol{Z}))\right] ,$$

which justifies our claim (S5.47).

For the term (S5.48). From Theorem 1 in Morikawa and Kim (2016), the efficient influence function of γ_0 is

$$\boldsymbol{\psi}_{eff}(T,\boldsymbol{Z};\gamma_0) = \mathbb{E}\left[\boldsymbol{S}_1(T,\boldsymbol{Z};\gamma_0)^{\otimes 2}\right]^{-1}\cdot \boldsymbol{S}_1(T,\boldsymbol{Z};\gamma_0) = -\mathbb{E}\left[\frac{\partial}{\partial\gamma}\boldsymbol{S}_1(T,\boldsymbol{Z};\gamma_0)\right]^{-1}\cdot \boldsymbol{S}_1(T,\boldsymbol{Z};\gamma_0) .$$

The efficient variance bound of γ_0 is

$$\boldsymbol{V}_{\gamma_0} = Var(\boldsymbol{\psi}_{eff}(T,\boldsymbol{Z};\gamma_0))$$

$$=\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}m(\boldsymbol{X})^{\otimes 2}\right]^{-1} \tag{S5.50}$$

$$=\mathbb{E}\left[\frac{m(\boldsymbol{X})}{\pi(\boldsymbol{Z};\gamma_0)}\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)^\intercal\right]^{-1}\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}m(\boldsymbol{X})^{\otimes 2}\right]\left\{\mathbb{E}\left[\frac{m(\boldsymbol{X})}{\pi(\boldsymbol{Z};\gamma_0)}\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)^\intercal\right]^{-1}\right\}^\intercal . \tag{S5.51}$$

Using (S5.42), the fact that $Var(\sqrt{N}(\bar{\gamma}-\gamma_0)) \to \boldsymbol{V}_{\gamma_0}$, and (S5.51), we can obtain that

$$\begin{aligned}
\text{L.H.S. of (S5.48)} &= \left\{\mathbb{E}\left[\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^\intercal\right]\cdot \boldsymbol{A}_{K\times K}^{-1}\cdot \mathbb{E}\left[u_K(\boldsymbol{X})\frac{\nabla\pi(\boldsymbol{Z};\gamma_0)^\intercal}{\pi(\boldsymbol{Z};\gamma_0)}\right]\right\}^{-1}\\
&= Var(\sqrt{N}(\bar{\gamma}-\gamma_0))+o(1)\\
&\to \boldsymbol{V}_{\gamma_0} = \mathbb{E}\left[\frac{m(\boldsymbol{X})}{\pi(\boldsymbol{Z};\gamma_0)}\nabla_\gamma\pi(\boldsymbol{Z};\gamma_0)^\intercal\right]^{-1}\mathbb{E}\left[\boldsymbol{S}_1(T,\boldsymbol{Z};\gamma_0)^{\otimes 2}\right]\left(\mathbb{E}\left[\frac{m(\boldsymbol{X})}{\pi(\boldsymbol{Z};\gamma_0)}\nabla_\gamma\pi(\boldsymbol{Z};\gamma_0)^\intercal\right]^{-1}\right)^\intercal ,
\end{aligned}$$

which justifies our claim (S5.48).

Combining (S5.45), (S5.46), (S5.47), (S5.48) and the definition of $\boldsymbol{\kappa}$ in (S1.5), we can obtain that

$$\begin{aligned}
&Var(\sqrt{N}(\bar{\theta}-\theta_0))\\
\to&\mathbb{E}\left[\left(\theta_0-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right]-\mathbb{E}\left[\left\{\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)R(\boldsymbol{X})\right\}^2\right]+\boldsymbol{\kappa}^\intercal\mathbb{E}[\boldsymbol{S}_1(T,\boldsymbol{Z};\gamma_0)^{\otimes 2}]\boldsymbol{\kappa} .
\end{aligned} \tag{S5.52}$$

To complete the proof, we remain to verify

$$\text{(S5.52)} = Var(S_2(T,\boldsymbol{Z};\gamma_0,\theta_0)-\boldsymbol{\kappa}^\intercal\boldsymbol{S}_1(T,\boldsymbol{Z};\gamma_0)) .$$

Note that

$$\begin{aligned}
&Var\left(S_2(T,\boldsymbol{Z};\gamma_0,\theta_0)-\boldsymbol{\kappa}^\intercal\boldsymbol{S}_1(T,\boldsymbol{Z};\gamma_0)\right)\\
=&\mathbb{E}\left[S_2(T,\boldsymbol{Z};\gamma_0,\theta_0)^2\right]-2\boldsymbol{\kappa}^\intercal\mathbb{E}\left[\boldsymbol{S}_1(T,\boldsymbol{Z};\gamma_0)S_2(T,\boldsymbol{Z};\gamma_0,\theta_0)\right]+\boldsymbol{\kappa}^\intercal\mathbb{E}\left[\boldsymbol{S}_1(T,\boldsymbol{Z};\gamma_0)^{\otimes 2}\right]\boldsymbol{\kappa} .
\end{aligned}$$

Note that

$$\mathbb{E}\left[S_2(T,\boldsymbol{Z};\gamma_0,\theta_0)^2\right]$$

$$
\begin{aligned}
=&\mathbb{E}\left[\left(\theta_0-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right]-2\cdot\mathbb{E}\left[\left(\theta_0-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)R(\boldsymbol{X})\right]\\
&+\mathbb{E}\left[\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)^2R(\boldsymbol{X})^2\right]\\
=&\mathbb{E}\left[\left(\theta_0-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right]+2\cdot\mathbb{E}\left[\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)R(\boldsymbol{X})\right]\\
&+\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}R(\boldsymbol{X})^2\right]\\
=&\mathbb{E}\left[\left(\theta_0-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right]-2\cdot\mathbb{E}\left[U(\boldsymbol{Z})\left(\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}\right)R(\boldsymbol{X})\right]+\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}R(\boldsymbol{X})^2\right]\\
=&\mathbb{E}\left[\left(\theta_0-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right]-2\cdot\mathbb{E}\left[\left(\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}\right)R(\boldsymbol{X})^2\right]+\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}R(\boldsymbol{X})^2\right]\\
=&\mathbb{E}\left[\left(\theta_0-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right]-\mathbb{E}\left[\left(\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}\right)R(\boldsymbol{X})^2\right]
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left[\boldsymbol{S}_1(T,\boldsymbol{Z};\gamma_0)S_2(T,\boldsymbol{Z};\gamma_0,\theta_0)\right]\\
=&-\mathbb{E}\left[\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)m(\boldsymbol{X})\cdot\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right]-\mathbb{E}\left[\left(1-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}\right)^2m(\boldsymbol{X})R(\boldsymbol{X})\right]\\
=&\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}m(\boldsymbol{X})\cdot U(\boldsymbol{Z})\right]-\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}m(\boldsymbol{X})\cdot R(\boldsymbol{X})\right]\\
=&\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}m(\boldsymbol{X})\cdot U(\boldsymbol{Z})\right]-\mathbb{E}\left[\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}\middle|\boldsymbol{X}\right]m(\boldsymbol{X})\cdot R(\boldsymbol{X})\right]\\
=&\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}m(\boldsymbol{X})\cdot U(\boldsymbol{Z})\right]-\mathbb{E}\left[m(\boldsymbol{X})\cdot\mathbb{E}[O(\boldsymbol{Z})U(\boldsymbol{Z})|\boldsymbol{X}]\right]\\
=&\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}m(\boldsymbol{X})\cdot U(\boldsymbol{Z})\right]-\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}m(\boldsymbol{X})\cdot U(\boldsymbol{Z})\right]\\
=&\mathbf{0}\ .
\end{aligned}
$$

Thus, we can conclude (S5.52) $= Var(S_2(T,\boldsymbol{Z};\gamma_0,\theta_0)-\boldsymbol{\kappa}^\intercal\boldsymbol{S}_1(T,\boldsymbol{Z};\gamma_0))$.

The proof is completed.

S6 Proof of Theorem 4

From the equation (S4.31), in order to prove $\|\sqrt{N}(\hat{\gamma}-\bar{\gamma},\hat{\theta}-\bar{\theta})\| \xrightarrow{p} 0$, it is sufficient to show

$$\widehat{\boldsymbol{D}}_{(K+1)\times(K+1)} \xrightarrow{p} \boldsymbol{D}_{(K+1)\times(K+1)} \ . \tag{S6.53}$$

Note that

$$\hat{\boldsymbol{D}}_{(K+1)\times(K+1)} := \begin{pmatrix} N^{-1}\sum_{i=1}^{N}\frac{1-\pi(\boldsymbol{Z}_i;\check{\gamma})}{\pi(\boldsymbol{Z}_i;\check{\gamma})}u_K(\boldsymbol{X}_i)^{\otimes 2}, & N^{-1}\sum_{i=1}^{N}\frac{1-\pi(\boldsymbol{Z}_i;\check{\gamma})}{\pi(\boldsymbol{Z}_i;\check{\gamma})}u_K(\boldsymbol{X}_i)U(\boldsymbol{Z}_i) \\ N^{-1}\sum_{i=1}^{N}\frac{1-\pi(\boldsymbol{Z}_i;\check{\gamma})}{\pi(\boldsymbol{Z}_i;\check{\gamma})}u_K(\boldsymbol{X}_i)^{\intercal}U(\boldsymbol{Z}_i), & N^{-1}\sum_{i=1}^{N}\left(\check{\theta}-\frac{T}{\pi(\boldsymbol{Z}_i;\check{\gamma})}U(\boldsymbol{Z}_i)\right)^2 \end{pmatrix}$$

and

$$\boldsymbol{D}_{(K+1)\times(K+1)} := \begin{pmatrix} \mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^{\otimes 2}\right], & \mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})U(\boldsymbol{Z})\right] \\ \mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^{\intercal}U(\boldsymbol{Z})\right], & \mathbb{E}\left[\left(\theta_0-\frac{T}{\pi(\boldsymbol{Z};\gamma_0)}U(\boldsymbol{Z})\right)^2\right] \end{pmatrix}$$

For simplicity, we show that the upper left block of $\widehat{\boldsymbol{D}}_{(K+1)\times(K+1)}$ converges in probability to that of $\boldsymbol{D}_{(K+1)\times(K+1)}$, namely

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\frac{1-\pi(\boldsymbol{Z}_i;\check{\gamma})}{\pi(\boldsymbol{Z}_i;\check{\gamma})}u_K(\boldsymbol{X}_i)^{\otimes 2}-\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^{\otimes 2}\right]\right\| \xrightarrow{p} 0 \ , \tag{S6.54}$$

and similar argument can be applied to the other three blocks are also of convergence.

Using Mean Value Theorem, we can obtain that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1-\pi(\boldsymbol{Z}_i;\check{\gamma})}{\pi(\boldsymbol{Z}_i;\check{\gamma})}u_K(\boldsymbol{X}_i)^{\otimes 2}-\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^{\otimes 2}\right]$$

$$=\frac{1}{N}\sum_{i=1}^{N}\frac{1-\pi(\boldsymbol{Z}_i;\gamma_0)}{\pi(\boldsymbol{Z}_i;\gamma_0)}u_K(\boldsymbol{X}_i)^{\otimes 2}-\mathbb{E}\left[\frac{1-\pi(\boldsymbol{Z};\gamma_0)}{\pi(\boldsymbol{Z};\gamma_0)}u_K(\boldsymbol{X})^{\otimes 2}\right] \tag{S6.55}$$

$$-(\check{\gamma}-\gamma_0)^\top\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{\nabla_\gamma\pi(\boldsymbol{Z}_i;\gamma^*)}{\pi(\boldsymbol{Z}_i;\gamma^*)^2}u_K(\boldsymbol{X}_i)^{\otimes 2}\ , \tag{S6.56}$$

where γ^* lies on the line joining γ_0 and $\check{\gamma}$. By computing the second moments of (S6.55), and using Chebyshev's inequality and Assumption 7, we can claim that the term (S6.55) is of $o_p(1)$.

Consider the term (S6.56). From Assumption 6, we know that the funciton $\nabla_\gamma\pi(\boldsymbol{Z};\gamma)$ is uniformly bounded and the propensity score $\pi(\boldsymbol{Z};\gamma)$ are uniformly bounded away from zero, thus we can find a finte constant $C>0$ such that

$$|(\text{S6.56})|\leq\|\check{\gamma}-\gamma_0\|\cdot C\cdot\left\|\frac{1}{N}\sum_{i=1}^{N}u_K(\boldsymbol{X}_i)^{\otimes 2}\right\|\ .$$

Using Chebyshev's inequality, Inequality (S2.12), and Assumption 7, we can deduce that

$$\left\|\frac{1}{N}\sum_{i=1}^{N}u_K(\boldsymbol{X}_i)^{\otimes 2}-\mathbb{E}\left[u_K(\boldsymbol{X})^{\otimes 2}\right]\right\|=O_p\left(\sqrt{\frac{K^3}{N}}\right)=o_p(1)\ .$$

We also note that $\|\mathbb{E}\left[u_K(\boldsymbol{X})^{\otimes 2}\right]\|\leq\lambda_{\max}\left(\mathbb{E}\left[u_K(\boldsymbol{X})^{\otimes 2}\right]\right)\cdot\|I_{K\times K}\|=O(\sqrt{K})$. Therefore, in light of Theorem 1 and Assumption 7 we can deduce that

$$|(\text{S6.56})|\leq O_p(N^{-1/2})\cdot C\cdot O(\sqrt{K})=o_p(1)\ .$$

Since the terms (S6.55) and (S6.56) are all of $o_p(1)$, we can justify (S6.54).

S7 Proof of Theorem 5

Note that

$$\widehat{\boldsymbol{V}}_K = \left\{ \widehat{\boldsymbol{B}}^\top_{(K+1)\times(p+1)} \widehat{\boldsymbol{D}}^{-1}_{(K+1)\times(K+1)} \widehat{\boldsymbol{B}}_{(K+1)\times(p+1)} \right\}^{-1} ,$$

$$\boldsymbol{V}_K = \left\{ \boldsymbol{B}^\top_{(K+1)\times(p+1)} \boldsymbol{D}^{-1}_{(K+1)\times(K+1)} \boldsymbol{B}_{(K+1)\times(p+1)} \right\}^{-1} ,$$

and $\boldsymbol{V}_K \xrightarrow{p} \boldsymbol{V}_{eff}$. From (S6.53), we know that $\widehat{\boldsymbol{D}}_{(K+1)\times(K+1)} \xrightarrow{p} \boldsymbol{D}_{(K+1)\times(K+1)}$. Therefore, to prove the consistency result $\widehat{\boldsymbol{V}}_K \xrightarrow{p} \boldsymbol{V}_{eff}$, it suffices to show $\widehat{\boldsymbol{B}}_{(K+1)\times(p+1)} \xrightarrow{p} \boldsymbol{B}_{(K+1)\times(p+1)}$.

We show that the upper left block of $\widehat{\boldsymbol{B}}_{(K+1)\times(p+1)}$ converges in probability to that of $\boldsymbol{B}_{(K+1)\times(p+1)}$, namely,

$$\frac{1}{N}\sum_{i=1}^{N} u_K(\boldsymbol{X}_i)\frac{\nabla_\gamma \pi(\boldsymbol{Z}_i;\hat{\gamma})^\top}{\pi(\boldsymbol{Z}_i;\hat{\gamma})^2} \xrightarrow{P} \mathbb{E}\left[u_K(\boldsymbol{X})\frac{\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)^\top}{\pi(\boldsymbol{Z};\gamma_0)}\right] , \tag{S7.57}$$

and similar arguments can be applied to show that the other three blocks are also of convergence. Using Mean Value Theorem, we can have

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N} u_K(\boldsymbol{X}_i)\frac{\nabla_\gamma \pi(\boldsymbol{Z}_i;\hat{\gamma})^\top}{\pi(\boldsymbol{Z}_i;\hat{\gamma})} \\
=&\frac{1}{N}\sum_{i=1}^{N} u_K(\boldsymbol{X}_i)\frac{\nabla_\gamma \pi(\boldsymbol{Z}_i;\gamma_0)^\top}{\pi(\boldsymbol{Z}_i;\gamma_0)} && \text{(S7.58)}\\
&- \frac{1}{N}\sum_{i=1}^{N} u_K(\boldsymbol{X}_i)\frac{\nabla_\gamma \pi(\boldsymbol{Z}_i;\gamma^{**})^\top}{\pi(\boldsymbol{Z}_i;\gamma^{**})^2}\nabla_\gamma \pi(\boldsymbol{Z}_i;\gamma^{**})^\top(\hat{\gamma}-\gamma_0) && \text{(S7.59)}\\
&+ \frac{1}{N}\sum_{i=1}^{N} u_K(\boldsymbol{X}_i)\frac{1}{\pi(\boldsymbol{Z}_i;\gamma^{**})}\cdot \nabla^2_\gamma \pi(\boldsymbol{Z}_i;\gamma^{**})(\hat{\gamma}-\gamma_0)\cdot , && \text{(S7.60)}
\end{aligned}$$

where γ^{**} lies on the line joining $\hat{\gamma}$ and γ_0. Consider the term (S7.58), we have

$$\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N} u_K(\boldsymbol{X}_i)\frac{\nabla_\gamma \pi(\boldsymbol{Z}_i;\gamma_0)^\top}{\pi(\boldsymbol{Z}_i;\gamma_0)} - \mathbb{E}\left[u_K(\boldsymbol{X})\frac{\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)^\top}{\pi(\boldsymbol{Z};\gamma_0)}\right]\right\|^2\right] \\
\leq&\frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}\left[\left\|u_K(\boldsymbol{X}_i)\frac{\nabla_\gamma \pi(\boldsymbol{Z}_i;\gamma_0)^\top}{\pi(\boldsymbol{Z}_i;\gamma_0)}\right\|^2\right] \\
\leq&\frac{1}{N}\mathbb{E}\left[\frac{\|\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)\|^2}{\pi(\boldsymbol{Z};\gamma_0)^2}\cdot\|u_K(\boldsymbol{X})\|^2\right] \leq O(1)\cdot\frac{1}{N}\mathbb{E}\left[\|u_K(\boldsymbol{X})\|^2\right] = O(K/N) = o(1)\ ,
\end{aligned}$$

where the second equality holds because $\frac{\|\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)\|^2}{\pi(\boldsymbol{Z};\gamma_0)}$ is uniformly bounded, while the last equality holds because of Assumption 7. Then in light of Markov's inequality, we can have

$$\left\|\frac{1}{N}\sum_{i=1}^{N} u_K(\boldsymbol{X}_i)\frac{\nabla_\gamma \pi(\boldsymbol{Z}_i;\gamma_0)^\top}{\pi(\boldsymbol{Z}_i;\gamma_0)} - \mathbb{E}\left[u_K(\boldsymbol{X})\frac{\nabla_\gamma \pi(\boldsymbol{Z};\gamma_0)^\top}{\pi(\boldsymbol{Z};\gamma_0)}\right]\right\| = o_p(1)\ .$$

For the terms (S7.59) and (S7.60), by using a similar argument of showing (S6.56) $= o_p(1)$, we can obtain that both (S7.59) and (S7.60) are of $o_p(1)$. Therefore, we can justify the validity of (S7.57). Finally, we can claim our consistency result $\widehat{\boldsymbol{V}}_K \xrightarrow{p} \boldsymbol{V}_{eff}$.

Bibliography

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models, *Handbook of econometrics* **6**: 5549–5632.

Eicker, F. (1966). A multivariate central limit theorem for random linear vector forms, *The Annals of Mathematical Statistics* pp. 1825–1828.

Horowitz, J. L. (2012). *Semiparametric methods in econometrics*, Vol. 131, Springer Science & Business Media.

Morikawa, K. and Kim, J. K. (2016). Semiparametric adaptive estimation with nonignorable nonresponse data, *arXiv preprint arXiv:1612.09207* .

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity, *Econometrica: Journal of the Econometric Society* pp. 1161–1167.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators, *Journal of econometrics* **79**(1): 147–168.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression, *Econometrica: Journal of the Econometric Society* pp. 931–954.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression, *The annals of statistics* pp. 1040–1053.

Stone, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation, *The Annals of Statistics* pp. 118–171.