Supplemental Material for "*A Unified Framework for Efficient Estimation of General Treatment Models*"

Chunrong Ai[*]– Chinese University of Hong Kong, Shenzhen

Oliver Linton[†]– University of Cambridge

Kaiji Motegi[‡]– Kobe University

Zheng Zhang[§]– Renmin University of China

This draft: August 24, 2020

[*]School of Management and Economics, Chinese University of Hong Kong, Shenzhen. E-mail: chunrongai@cuhk.edu.cn

[†]Faculty of Economics, University of Cambridge. E-mail: obl20@cam.ac.uk

[‡]Graduate School of Economics, Kobe University. E-mail: motegi@econ.kobe-u.ac.jp

[§]Institute of Statistics and Big Data, Renmin University of China. E-mail: zhengzhang@ruc.edu.cn

Contents

1 Assumptions

Assumption 1.1 (*Unconfounded Treatment Assignment*) *For all $t \in \mathcal{T}$, given $\boldsymbol{X}$, T is independent of $Y^*(t)$, i.e., $Y^*(t) \perp T | \boldsymbol{X}$, for all $t \in \mathcal{T}$.*

Assumption 1.2 *(i) The support $\mathcal{X}$ of $\boldsymbol{X}$ is a compact subset of $\mathbb{R}^r$. The support $\mathcal{T}$ of the treatment variable T is a compact subset of $\mathbb{R}$. (ii) There exist two positive constants η_1 and η_2 such that*

$$0 < \eta_1 \leq \pi_0(t, \boldsymbol{x}) \leq \eta_2 < \infty \ , \ \forall (t, \boldsymbol{x}) \in \mathcal{T} \times \mathcal{X} \ .$$

Assumption 1.3 *There exist $\Lambda_{K_1 \times K_2} \in \mathbb{R}^{K_1 \times K_2}$ and a positive constant $\alpha > 0$ such that*

$$\sup_{(t,\boldsymbol{x}) \in \mathcal{T} \times \mathcal{X}} \left| (\rho'^{-1} \left(\pi_0(t, \boldsymbol{x}) \right) - u_{K_1}(t)^\top \Lambda_{K_1 \times K_2} v_{K_2}(\boldsymbol{x}) \right| = O(K^{-\alpha}).$$

Assumption 1.4 *(i) For every K_1 and K_2, the smallest eigenvalues of $\mathbb{E}\left[u_{K_1}(T) u_{K_1}(T)^\top\right]$ and $\mathbb{E}\left[v_{K_2}(\boldsymbol{X}) v_{K_2}(\boldsymbol{X})^\top\right]$ are bounded away from zero uniformly in K_1 and K_2. (ii) There are two sequences of constants $\zeta_1(K_1)$ and $\zeta_2(K_2)$ satisfying $\sup_{t \in \mathcal{T}} \|u_{K_1}(t)\| \leq \zeta_1(K_1)$ and $\sup_{\boldsymbol{x} \in \mathcal{X}} \|v_{K_2}(\boldsymbol{x})\| \leq \zeta_2(K_2)$, $K = K_1(N) K_2(N)$ and $\zeta(K) := \zeta_1(K_1)\zeta_2(K_2)$, such that $\zeta(K) K^{-\alpha} \to 0$ and $\zeta(K)\sqrt{K/N} \to 0$ as $N \to \infty$.*

Assumption 1.5 *(i) The parameter space $\Theta \subset \mathbb{R}^p$ is a compact set and the true parameter $\boldsymbol{\beta}^*$ is in the interior of Θ , where $p \in \mathbb{N}$. (ii) $L\left(Y - g(T; \boldsymbol{\beta})\right)$ is continuous in $\boldsymbol{\beta}$, $\sup_{\boldsymbol{\beta} \in \Theta} \mathbb{E}[|L(Y - g(T; \boldsymbol{\beta}))|^2] < \infty$ and $\mathbb{E}[\sup_{\boldsymbol{\beta} \in \Theta} |L\left(Y - g(T; \boldsymbol{\beta})\right)|] < \infty$.*

Assumption 1.6

(i) The loss function $L(v)$ is differentiable almost everywhere, $g(t; \boldsymbol{\beta})$ is twice continuously differentiable in $\boldsymbol{\beta} \in \Theta$ and we denote its first derivative by $m(t; \boldsymbol{\beta}) := \nabla_\beta g(t; \boldsymbol{\beta})$;

(ii) $\mathbb{E}\left[\pi_0(T, \boldsymbol{X}) L'(Y - g(T; \boldsymbol{\beta})) m(T; \boldsymbol{\beta})\right]$ is differentiable with respect to $\boldsymbol{\beta}$ and $H_0 := -\nabla_\beta \mathbb{E}[\pi_0(T, \boldsymbol{X}) L'(Y - g(T; \boldsymbol{\beta})) m(T;\ \boldsymbol{\beta})]\Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}_0}$ is nonsingular;

(iii) $\varepsilon(t, \boldsymbol{x}; \boldsymbol{\beta}_0) := \mathbb{E}[L'(Y - g(T; \boldsymbol{\beta}_0)) | T = t, \boldsymbol{X} = \boldsymbol{x}]$ is continuously differentiable in $(t, \boldsymbol{x})$;

(iv) Suppose that $N^{-1} \sum_{i=1}^N \hat{\pi}_K(T_i, \boldsymbol{X}_i) L'\left(Y_i - g(T_i; \hat{\boldsymbol{\beta}})\right) m(T_i; \hat{\boldsymbol{\beta}}) = o_p(N^{-1/2})$ holds with probability approaching one.

Assumption 1.7 *(i)* $\mathbb{E}\left[\sup_{\boldsymbol{\beta}\in\Theta}|L'(Y-g(T;\boldsymbol{\beta}))|^{2+\delta}\right]<\infty$ *for some* $\delta>0$*; (ii) The function class* $\{L'(y-g(t;\boldsymbol{\beta})):\boldsymbol{\beta}\in\Theta\}$ *satisfies:*

$$\mathbb{E}\left[\sup_{\boldsymbol{\beta}_1:\|\boldsymbol{\beta}_1-\boldsymbol{\beta}\|<\delta}\left|L'(Y-g(T;\boldsymbol{\beta}_1))-L'(Y-g(T;\boldsymbol{\beta}))\right|^2\right]^{1/2}\leq a\cdot\delta^b$$

for any $\boldsymbol{\beta}\in\Theta$ *and any small* $\delta>0$ *and for some finite positive constants* a *and* b*.*

Assumption 1.8 $\zeta(K)\sqrt{K^2/N}\to 0$ *and* $\sqrt{N}K^{-\alpha}\to 0$*.*

2 Efficiency Bound

2.1 Proof of Theorem 1

Without loss of generality, we only consider the distribution of $(T,\boldsymbol{X},Y)$ to be absolutely continuous with respect to Lebesgue measure, i.e., there exists a density function $f_{T,X,Y}(t,\boldsymbol{x},y)$ such that $dF_{T,X,Y}(t,\boldsymbol{x},y)=f_{T,X,Y}(t,\boldsymbol{x},y)dtd\boldsymbol{x}dy$. For discrete cases, the proof can be established by using a similar argument.

We follow the approach of Bickel, Klaassen, Ritov, and Wellner (1993, Section 3.3) to derive the variance bound of $\boldsymbol{\beta}^*$, see also Tchetgen Tchetgen and Shpitser (2012). Let $\left\{f^{\alpha}_{Y,T,X}(y,t,\boldsymbol{x})\right\}_{\alpha\in\mathbb{R}}$ denote a one dimensional regular parametric submodel with $f^{\alpha=0}_{Y,T,X}(y,t,\boldsymbol{x})=f_{Y,T,X}(y,t,\boldsymbol{x})$. By definition, $\boldsymbol{\beta}^*$ solves following equation:

$$\int_{\mathcal{T}}\mathbb{E}\left[m(t;\boldsymbol{\beta}^*)L'\left(Y^*(t)-g\left(t;\boldsymbol{\beta}^*\right)\right)\right]f_T(t)dt=0\ . \tag{1}$$

By Assumption 1.1, (1) is equivalent to

$$\int_{\mathcal{T}}\int_{\mathcal{X}}\mathbb{E}\left[m(T;\boldsymbol{\beta}^*)L'\left(Y-g\left(T;\boldsymbol{\beta}^*\right)\right)|T=t,\boldsymbol{X}=\boldsymbol{x}\right]f_X(\boldsymbol{x})f_T(t)d\boldsymbol{x}dt=0\ .$$

Therefore, the parameter $\boldsymbol{\beta}(\alpha)$ induced by the submodel $f^{\alpha}_{Y,T,X}(y,t,\boldsymbol{x})$ satisfies:

$$\int_{\mathcal{T}}\int_{\mathcal{X}}m(t;\boldsymbol{\beta}(\alpha))\cdot\mathbb{E}^{\alpha}\left[L'\left(Y-g\left(t;\boldsymbol{\beta}(\alpha)\right)\right)|T=t,\boldsymbol{X}=\boldsymbol{x}\right]f^{\alpha}_T(t)f^{\alpha}_X(\boldsymbol{x})d\boldsymbol{x}dt=0\ , \tag{2}$$

where $\mathbb{E}^{\alpha}\left[\cdot|T=t,\boldsymbol{X}=\boldsymbol{x}\right]$ denotes taking expectation with respect to the submodel $f^{\alpha}_{Y|T,X}(\cdot|t,\boldsymbol{x})$.

Differentiating both sides of (2) with respect to α, evaluating at $\alpha = 0$ and using the condition $Y^*(t) \perp T|\boldsymbol{X}$, we can deduce that

$$
\begin{aligned}
0 =& \int_{\mathcal{T}}\int_{\mathcal{X}} \frac{\partial}{\partial\alpha}\bigg|_{\alpha=0} \left\{m(t;\boldsymbol{\beta}(\alpha))\mathbb{E}^{\alpha}\left[L'(Y-g(t;\boldsymbol{\beta}(\alpha)))|T=t,\boldsymbol{X}=\boldsymbol{x}\right]f_T^{\alpha}(t)f_X^{\alpha}(\boldsymbol{x})\right\}d\boldsymbol{x}dt \\
=& \int_{\mathcal{T}}\int_{\mathcal{X}} \mathbb{E}\left[L'(Y-g(t;\boldsymbol{\beta}^*))|T=t,\boldsymbol{X}=\boldsymbol{x}\right]f_T(t)f_X(\boldsymbol{x})\nabla_{\boldsymbol{\beta}}m(t;\boldsymbol{\beta}^*)d\boldsymbol{x}dt\cdot\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}\boldsymbol{\beta}(\alpha) \\
&+\int_{\mathcal{X}\times\mathcal{T}} \mathbb{E}[L'(Y-g(t;\boldsymbol{\beta}^*))|T=t,\boldsymbol{X}=\boldsymbol{x}]m(t;\boldsymbol{\beta}^*)\cdot\frac{\partial}{\partial\alpha}f_X^{\alpha}(\boldsymbol{x})\bigg|_{\alpha=0}f_T(t)d\boldsymbol{x}dt \\
&+\int_{\mathcal{Y}\times\mathcal{X}\times\mathcal{T}} m(t;\boldsymbol{\beta}^*)L'(y-g(t;\boldsymbol{\beta}^*))\cdot\frac{\partial}{\partial\alpha}f_{Y|T,X}^{\alpha}(y|t,\boldsymbol{x})\bigg|_{\alpha=0}f_X(\boldsymbol{x})f_T(t)dyd\boldsymbol{x}dt \\
&+\int_{\mathcal{X}\times\mathcal{T}} m(t;\boldsymbol{\beta}^*)\cdot\nabla_{\beta}\mathbb{E}[L'(Y^*(t)-g(t;\boldsymbol{\beta}))|T=t,\boldsymbol{X}=\boldsymbol{x}]\bigg|_{\beta=\boldsymbol{\beta}^*}\cdot\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}\boldsymbol{\beta}(\alpha)\cdot f_T(t)f_X(\boldsymbol{x})d\boldsymbol{x}dt \\
&+\int_{\mathcal{X}\times\mathcal{T}} \mathbb{E}[L'(Y-g(t;\boldsymbol{\beta}^*))|T=t,\boldsymbol{X}=\boldsymbol{x}]m(t;\boldsymbol{\beta}^*)\cdot\frac{\partial}{\partial\alpha}f_T^{\alpha}(t)\bigg|_{\alpha=0}f_X(\boldsymbol{x})d\boldsymbol{x}dt \\
=& \int_{\mathcal{T}}\int_{\mathcal{X}} \mathbb{E}\left[L'(Y^*(t)-g(t;\boldsymbol{\beta}^*))|\boldsymbol{X}=\boldsymbol{x}\right]f_T(t)f_X(\boldsymbol{x})\nabla_{\boldsymbol{\beta}}m(t;\boldsymbol{\beta}^*)d\boldsymbol{x}dt\cdot\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}\boldsymbol{\beta}(\alpha) \\
&+\int_{\mathcal{X}\times\mathcal{T}} \mathbb{E}[L'(Y-g(t;\boldsymbol{\beta}^*))|T=t,\boldsymbol{X}=\boldsymbol{x}]m(t;\boldsymbol{\beta}^*)\cdot\frac{\partial}{\partial\alpha}f_X^{\alpha}(\boldsymbol{x})\bigg|_{\alpha=0}f_T(t)d\boldsymbol{x}dt \\
&+\int_{\mathcal{Y}\times\mathcal{X}\times\mathcal{T}} m(t;\boldsymbol{\beta}^*)L'(y-g(t;\boldsymbol{\beta}^*))\cdot\frac{\partial}{\partial\alpha}f_{Y|T,X}^{\alpha}(y|t,\boldsymbol{x})\bigg|_{\alpha=0}f_X(\boldsymbol{x})f_T(t)dyd\boldsymbol{x}dt \\
&+\int_{\mathcal{X}\times\mathcal{T}} m(t;\boldsymbol{\beta}^*)\cdot\nabla_{\beta}\mathbb{E}[L'(Y^*(t)-g(t;\boldsymbol{\beta}))|\boldsymbol{X}=\boldsymbol{x}]\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\cdot f_T(t)f_X(\boldsymbol{x})d\boldsymbol{x}dt\cdot\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}\boldsymbol{\beta}(\alpha) \\
&+\int_{\mathcal{X}\times\mathcal{T}} \mathbb{E}[L'(Y-g(t;\boldsymbol{\beta}^*))|T=t,\boldsymbol{X}=\boldsymbol{x}]m(t;\boldsymbol{\beta}^*)\cdot\frac{\partial}{\partial\alpha}f_T^{\alpha}(t)\bigg|_{\alpha=0}f_X(\boldsymbol{x})d\boldsymbol{x}dt \\
=& \int_{\mathcal{T}} \mathbb{E}[L'(Y^*(t)-g(t;\boldsymbol{\beta}^*))]\cdot f_T(t)\nabla_{\boldsymbol{\beta}}m(t;\boldsymbol{\beta}^*)dt\cdot\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}\boldsymbol{\beta}(\alpha) \\
&+\int_{\mathcal{X}\times\mathcal{T}} \mathbb{E}[L'(Y-g(t;\boldsymbol{\beta}^*))|T=t,\boldsymbol{X}=\boldsymbol{x}]m(t;\boldsymbol{\beta}^*)\cdot\frac{\partial}{\partial\alpha}f_X^{\alpha}(\boldsymbol{x})\bigg|_{\alpha=0}f_T(t)d\boldsymbol{x}dt \\
&+\int_{\mathcal{Y}\times\mathcal{X}\times\mathcal{T}} m(t;\boldsymbol{\beta}^*)\cdot L'(y-g(t;\boldsymbol{\beta}^*))\cdot\frac{\partial}{\partial\alpha}f_{Y|T,X}^{\alpha}(y|t,\boldsymbol{x})\bigg|_{\alpha=0}f_X(\boldsymbol{x})f_T(t)dyd\boldsymbol{x}dt \\
&+\int_{\mathcal{T}} \nabla_{\beta}\mathbb{E}[L'(Y^*(t)-g(t;\boldsymbol{\beta}))]\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}m(t;\boldsymbol{\beta}^*)\cdot f_T(t)dt\cdot\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}\boldsymbol{\beta}(\alpha) \\
&+\int_{\mathcal{X}\times\mathcal{T}} \mathbb{E}[L'(Y-g(t;\boldsymbol{\beta}^*))|T=t,\boldsymbol{X}=\boldsymbol{x}]m(t;\boldsymbol{\beta}^*)\cdot\frac{\partial}{\partial\alpha}f_T^{\alpha}(t)\bigg|_{\alpha=0}f_X(\boldsymbol{x})d\boldsymbol{x}dt \\
=&\nabla_{\beta}\left\{\int_{\mathcal{T}} \mathbb{E}[L'(Y^*(t)-g(t;\boldsymbol{\beta}))]\cdot m(t;\boldsymbol{\beta})f_T(t)dt\right\}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\cdot\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}\boldsymbol{\beta}(\alpha)
\end{aligned}
$$

$$+\int_{\mathcal{X}\times\mathcal{T}}\mathbb{E}[L'(Y-g(t;\boldsymbol{\beta}^*))|T=t,\boldsymbol{X}=\boldsymbol{x}]m(t;\boldsymbol{\beta}^*)\cdot\frac{\partial}{\partial\alpha}f_X^\alpha(\boldsymbol{x})\bigg|_{\alpha=0}f_T(t)d\boldsymbol{x}dt$$
$$+\int_{\mathcal{Y}\times\mathcal{X}\times\mathcal{T}}m(t;\boldsymbol{\beta}^*)\cdot L'(y-g(t;\boldsymbol{\beta}^*))\cdot\frac{\partial}{\partial\alpha}f_{Y|T,X}^\alpha(y|t,\boldsymbol{x})\bigg|_{\alpha=0}f_X(\boldsymbol{x})f_T(t)dyd\boldsymbol{x}dt$$
$$+\int_{\mathcal{X}\times\mathcal{T}}\mathbb{E}[L'(Y-g(t;\boldsymbol{\beta}^*))|T=t,\boldsymbol{X}=\boldsymbol{x}]m(t;\boldsymbol{\beta}^*)\cdot\frac{\partial}{\partial\alpha}f_T^\alpha(t)\bigg|_{\alpha=0}f_X(\boldsymbol{x})d\boldsymbol{x}dt.$$

Since $H_0=-\nabla_\beta\left\{\int_{\mathcal{T}}\mathbb{E}[L'(Y^*(t)-g(t;\boldsymbol{\beta}))]\cdot m(t;\boldsymbol{\beta})f_T(t)dt\right\}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}$ is invertible, we get

$$\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}\boldsymbol{\beta}(\alpha)=H_0^{-1}\cdot\left\{\int_{\mathcal{X}\times\mathcal{T}}\mathbb{E}[L'(Y-g(t;\boldsymbol{\beta}^*))|T=t,\boldsymbol{X}=\boldsymbol{x}]m(t;\boldsymbol{\beta}^*)\cdot\frac{\partial}{\partial\alpha}f_X^\alpha(\boldsymbol{x})\bigg|_{\alpha=0}f_T(t)d\boldsymbol{x}dt\right.$$
$$+\int_{\mathcal{Y}\times\mathcal{X}\times\mathcal{T}}m(t;\boldsymbol{\beta}^*)\cdot L'(y-g(t;\boldsymbol{\beta}^*))\cdot\frac{\partial}{\partial\alpha}f_{Y|T,X}^\alpha(y|t,\boldsymbol{x})\bigg|_{\alpha=0}f_X(\boldsymbol{x})f_T(t)dyd\boldsymbol{x}dt$$
$$\left.+\int_{\mathcal{X}\times\mathcal{T}}\mathbb{E}[L'(Y-g(t;\boldsymbol{\beta}^*))|T=t,\boldsymbol{X}=\boldsymbol{x}]m(t;\boldsymbol{\beta}^*)\cdot\frac{\partial}{\partial\alpha}f_T^\alpha(t)\bigg|_{\alpha=0}f_X(\boldsymbol{x})d\boldsymbol{x}dt\right\}.$$

The efficient influence function of $\boldsymbol{\beta}^*$, denoted by $S_{eff}(Y,T,\boldsymbol{X};\boldsymbol{\beta}^*)$, is a unique function satisfying the following equation:

$$\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}\boldsymbol{\beta}(\alpha)=\mathbb{E}\left[S_{eff}(Y,T,\boldsymbol{X};\boldsymbol{\beta}^*)\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}\log f_{Y,X,T}^\alpha(Y,\boldsymbol{X},T)\right]. \tag{3}$$

Therefore, to justify our theorem, it suffices to substitute $S_{eff}(Y,T,\boldsymbol{X};\boldsymbol{\beta}^*)=H_0^{-1}\psi(Y,T,\boldsymbol{X};\boldsymbol{\beta}^*)$ into (3) and check the validity. Note that

$$\mathbb{E}\left[S_{eff}(Y,T,\boldsymbol{X};\boldsymbol{\beta}^*)\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}\log f_{Y,X,T}^\alpha(Y,\boldsymbol{X},T)\right]$$
$$=H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}\times\mathcal{Y}}\psi(y,t,\boldsymbol{x};\boldsymbol{\beta}^*)\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}f_{Y|X,T}^\alpha(y|\boldsymbol{x},t)f_{T,X}(t,\boldsymbol{x})dyd\boldsymbol{x}dt \tag{4}$$
$$+H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}\times\mathcal{Y}}\psi(y,t,\boldsymbol{x};\boldsymbol{\beta}^*)f_{Y|X,T}(y|\boldsymbol{x},t)\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}f_{T|X}^\alpha(t|\boldsymbol{x})f_X(\boldsymbol{x})dyd\boldsymbol{x}dt \tag{5}$$
$$+H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}\times\mathcal{Y}}\psi(y,t,\boldsymbol{x};\boldsymbol{\beta}^*)f_{Y|X,T}(y|\boldsymbol{x},t)f_{T|X}(t|\boldsymbol{x})\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}f_X^\alpha(\boldsymbol{x})dyd\boldsymbol{x}dt. \tag{6}$$

For the term (4), we have

$$(4)=H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}\times\mathcal{Y}}\left\{\frac{f_T(t)}{f_{T|X}(t|\boldsymbol{x})}m(t;\boldsymbol{\beta}^*)\cdot L'(y-g(t;\boldsymbol{\beta}^*))-\frac{f_T(t)}{f_{T|X}(t|\boldsymbol{x})}m(t;\boldsymbol{\beta}^*)\cdot\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\right.$$

$$
\begin{aligned}
&+\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|\boldsymbol{X}=\boldsymbol{x}\right]+\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|T=t\right]\Big\} \\
&\times \frac{\partial}{\partial\alpha}\Big|_{\alpha=0} f^{\alpha}_{Y|X,T}(y|\boldsymbol{x},t)f_{T,X}(t,\boldsymbol{x})dyd\boldsymbol{x}dt \\
=&H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}\times\mathcal{Y}}\frac{f_T(t)}{f_{T|X}(t|\boldsymbol{x})}m(t;\boldsymbol{\beta}^*)\cdot L'(y-g(t;\boldsymbol{\beta}^*))\cdot\frac{\partial}{\partial\alpha}\Big|_{\alpha=0} f^{\alpha}_{Y|X,T}(y|\boldsymbol{x},t)f_{T,X}(t,\boldsymbol{x})dyd\boldsymbol{x}dt \\
=&H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}\times\mathcal{Y}}m(t;\boldsymbol{\beta}^*)\cdot L'(y-g(t;\boldsymbol{\beta}^*))\cdot\frac{\partial}{\partial\alpha}\Big|_{\alpha=0} f^{\alpha}_{Y|X,T}(y|\boldsymbol{x},t)f_T(t)f_X(\boldsymbol{x})dyd\boldsymbol{x}dt.
\end{aligned}
$$

For the term (5), we have

$$
\begin{aligned}
(5)=&H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}\times\mathcal{Y}}\Big\{\frac{f_T(t)}{f_{T|X}(t|\boldsymbol{x})}m(t;\boldsymbol{\beta}^*)\cdot L'(y-g(t;\boldsymbol{\beta}^*))-\frac{f_T(t)}{f_{T|X}(t|\boldsymbol{x})}m(t;\boldsymbol{\beta}^*)\cdot\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*) \\
&+\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|\boldsymbol{X}=\boldsymbol{x}\right]+\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|T=t\right]\Big\} \\
&\times f_{Y|X,T}(y|\boldsymbol{x},t)\frac{\partial}{\partial\alpha}\Big|_{\alpha=0} f^{\alpha}_{T|X}(t|\boldsymbol{x})f_X(\boldsymbol{x})dyd\boldsymbol{x}dt \\
=&H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}}\Big\{\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|\boldsymbol{X}=\boldsymbol{x}\right]+\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|T=t\right]\Big\} \\
&\cdot\frac{\partial}{\partial\alpha}\Big|_{\alpha=0} f^{\alpha}_{T|X}(t|\boldsymbol{x})f_X(\boldsymbol{x})d\boldsymbol{x}dt \\
=&H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}}\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|T=t\right]\cdot\frac{\partial}{\partial\alpha}\Big|_{\alpha=0} f^{\alpha}_{T|X}(t|\boldsymbol{x})f_X(\boldsymbol{x})d\boldsymbol{x}dt \\
=&H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}}\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|T=t\right]\cdot\frac{\partial}{\partial\alpha}\Big|_{\alpha=0} f^{\alpha}_{T}(t)dt \\
=&H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}}\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\frac{f_T(t)}{f_{T|X}(t|\boldsymbol{x})}m(t;\boldsymbol{\beta}^*)\cdot\frac{\partial}{\partial\alpha}\Big|_{\alpha=0} f^{\alpha}_{T}(t)\cdot f_{X|T}(\boldsymbol{x}|t)d\boldsymbol{x}dt \\
=&H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}}\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)m(t;\boldsymbol{\beta}^*)\cdot\frac{\partial}{\partial\alpha}\Big|_{\alpha=0} f^{\alpha}_{T}(t)\cdot f_X(\boldsymbol{x})d\boldsymbol{x}dt,
\end{aligned}
$$

where the first equality holds in accordance with the definition of $\int_{\mathcal{Y}} L'(y-g(t;\boldsymbol{\beta}^*))f_{Y|X,T}(y|\boldsymbol{x},t)dy =: \varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)$.

For the term (6), we have

$$
\begin{aligned}
(6)=&H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}\times\mathcal{Y}}\Big\{\frac{f_T(t)}{f_{T|X}(t|\boldsymbol{x})}m(t;\boldsymbol{\beta}^*)\cdot L'(y-g(t;\boldsymbol{\beta}^*))-\frac{f_T(t)}{f_{T|X}(t|\boldsymbol{x})}m(t;\boldsymbol{\beta}^*)\cdot\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*) \\
&+\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|\boldsymbol{X}=\boldsymbol{x}\right]+\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|T=t\right]\Big\}
\end{aligned}
$$

$$
\begin{aligned}
&\times f_{Y|X,T}(y|\boldsymbol{x},t)f_{T|X}(t|\boldsymbol{x})\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}f_X^{\alpha}(\boldsymbol{x})dyd\boldsymbol{x}dt\\
=&H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}}\Big\{\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|\boldsymbol{X}=\boldsymbol{x}\right]+\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|T=t\right]\Big\}\\
&\times f_{T|X}(t|\boldsymbol{x})\cdot\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}f_X^{\alpha}(\boldsymbol{x})d\boldsymbol{x}dt\\
=&H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}}\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|\boldsymbol{X}=\boldsymbol{x}\right]\cdot f_{T|X}(t|\boldsymbol{x})\cdot\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}f_X^{\alpha}(\boldsymbol{x})d\boldsymbol{x}dt\\
=&H_0^{-1}\int_{\mathcal{X}}\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|\boldsymbol{X}=\boldsymbol{x}\right]\cdot\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}f_X^{\alpha}(\boldsymbol{x})d\boldsymbol{x}\\
=&H_0^{-1}\int_{\mathcal{X}\times\mathcal{T}}\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)m(t;\boldsymbol{\beta}^*)\cdot f_T(t)\cdot\frac{\partial}{\partial\alpha}\bigg|_{\alpha=0}f_X^{\alpha}(\boldsymbol{x})d\boldsymbol{x}dt.
\end{aligned}
$$

We have proved (3) holds, hence S_{eff} is the efficient influence function of $\boldsymbol{\beta}^*$.

2.2 Particular Case I: Binary Average Treatment Effects

In this section, we show that when $T\in\{0,1\}$, $g(t;\boldsymbol{\beta})=\beta_0+\beta_1\cdot t$ and $L(v)=v^2$, our general efficiency bound derived in Theorem 1 reduces to the well-known efficiency bound for average treatment effects in Robins, Rotnitzky, and Zhao (1994) and Hahn (1998). In accordance with our identification condition, β_0^* and β_1^* are identified by minimizing the following loss function

$$
\sum_{t\in\{0,1\}}\mathbb{E}[(Y^*(t)-\beta_0-\beta_1\cdot t)^2]\cdot\mathbb{P}(T=t).
$$

The solutions are given by

$$
\beta_0^*=\mathbb{E}[Y^*(0)],\ \beta_1^*=\mathbb{E}[Y^*(1)-Y^*(0)].
$$

Here β_1^* is the average treatment effects.

Corollary 2.1 *Suppose $T\in\{0,1\}$, $L(v)=v^2$, $g(t;\boldsymbol{\beta})=\beta_0+\beta_1\cdot t$ and the conditions in Theorem 1 hold, the efficient influence functions of β_0^* and β_1^* given by Theorem 1 reduce to*

$$
\begin{aligned}
&S_{eff}(T,\boldsymbol{X},Y;\beta_0^*)=\phi_2(T,\mathbf{X},Y;\beta_0^*),\\
&S_{eff}(T,\boldsymbol{X},Y;\beta_1^*,\beta_0^*)=\phi_2(T,\mathbf{X},Y;\beta_0^*)-\phi_1(T,\mathbf{X},Y;\beta_1^*,\beta_0^*),
\end{aligned}
$$

where

$$\phi_1(T,\boldsymbol{X},Y;\beta_1^*,\beta_0^*) = \frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot Y^*(1) - \left\{\frac{T}{\mathbb{P}(T=1|\boldsymbol{X})} - 1\right\}\cdot \mathbb{E}[Y^*(1)|\boldsymbol{X}] - \beta_0^* - \beta_1^*,$$
$$\phi_2(T,\boldsymbol{X},Y;\beta_0^*) = \frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}\cdot Y^*(0) - \left\{\frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})} - 1\right\}\cdot \mathbb{E}[Y^*(0)|\boldsymbol{X}] - \beta_0^*,$$

and they are the same as the efficient influence functions given in Robins, Rotnitzky, and Zhao (1994) and Hahn (1998).

Proof. Using our notation, we have

$$\boldsymbol{\beta}^* = (\beta_0,\beta_1)^\top,\ g(t;\boldsymbol{\beta}^*) = \beta_0^* + \beta_1^*\cdot t,\quad m(t;\boldsymbol{\beta}^*) = \begin{bmatrix}1\\ t\end{bmatrix},\ H_0 = \mathbb{E}\left[m(T;\boldsymbol{\beta}^*)m(T;\boldsymbol{\beta}^*)^\top\right]\ ,$$
$$\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*) = T\cdot\{\mathbb{E}[Y^*(1)-Y^*(0)|\boldsymbol{X}] - \beta_1^*\} + \mathbb{E}[Y^*(0)|\boldsymbol{X}] - \beta_0^*,$$
$$\pi_0(T,\boldsymbol{X}) = \frac{T\cdot p + (1-T)\cdot q}{T\cdot\mathbb{P}(T=1|\boldsymbol{X}) + T\cdot\mathbb{P}(T=0|\boldsymbol{X})} = \frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot p + \frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}\cdot q,$$

where $p = \mathbb{P}(T=1)$ and $q = \mathbb{P}(T=0)$. In accordance with our Theorem 1, the efficient influence function of (β_0,β_1) is

$$H_0^{-1}\left\{\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\left\{Y - \mathbb{E}[Y|\boldsymbol{X},T]\right\} + \mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|\boldsymbol{X}\right]\right\}.$$

With some computation, we have

$$H_0^{-1} = \begin{bmatrix}1 & p\\ p & p\end{bmatrix}^{-1} = \frac{1}{pq}\cdot\begin{bmatrix}p & -p\\ -p & 1\end{bmatrix} = \begin{bmatrix}\frac{1}{q} & -\frac{1}{q}\\ -\frac{1}{q} & \frac{1}{pq}\end{bmatrix}. \tag{7}$$

and

$$\begin{aligned}
&\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\left\{Y - \mathbb{E}[Y|\boldsymbol{X},T]\right\}\\
=&\frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot p\cdot\begin{bmatrix}1\\ T\end{bmatrix}\cdot\left\{Y - T\cdot\mathbb{E}[Y^*(1)|\boldsymbol{X}] - (1-T)\cdot\mathbb{E}[Y^*(0)|\boldsymbol{X}]\right\}\\
&+\frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}\cdot q\cdot\begin{bmatrix}1\\ T\end{bmatrix}\cdot\left\{Y - T\cdot\mathbb{E}[Y^*(1)|\boldsymbol{X}] - (1-T)\cdot\mathbb{E}[Y^*(0)|\boldsymbol{X}]\right\}\\
=&\frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot p\cdot\begin{bmatrix}1\\ 1\end{bmatrix}\cdot\left\{Y^*(1) - \mathbb{E}[Y^*(1)|\boldsymbol{X}]\right\} + \frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}\cdot q\cdot\begin{bmatrix}1\\ 0\end{bmatrix}\cdot\left\{Y^*(0) - \mathbb{E}[Y^*(0)|\boldsymbol{X}]\right\}
\end{aligned}$$

$$
=\begin{bmatrix} \frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot\{Y^*(1)-\mathbb{E}[Y^*(1)|\boldsymbol{X}]\}\cdot p+\frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}\cdot\{Y^*(0)-\mathbb{E}[Y^*(0)|\boldsymbol{X}]\}\cdot q \\ \frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot\{Y^*(1)-\mathbb{E}[Y^*(1)|\boldsymbol{X}]\}\cdot p \end{bmatrix} \tag{8}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|\boldsymbol{X}\right] \\
=&\mathbb{E}\left[\left(T\cdot\{\mathbb{E}[Y^*(1)-Y^*(0)|\boldsymbol{X}]-\beta_1^*\}+\mathbb{E}[Y^*(0)|\boldsymbol{X}]-\beta_0^*\right)\cdot\frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot p\cdot\begin{bmatrix}1\\T\end{bmatrix}\middle|\boldsymbol{X}\right] \\
&+\mathbb{E}\left[\left(T\cdot\{\mathbb{E}[Y^*(1)-Y^*(0)|\boldsymbol{X}]-\beta_1^*\}+\mathbb{E}[Y^*(0)|\boldsymbol{X}]-\beta_0^*\right)\cdot\frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}\cdot q\cdot\begin{bmatrix}1\\T\end{bmatrix}\middle|\boldsymbol{X}\right] \\
=&\mathbb{E}\left[\left(\mathbb{E}[Y^*(1)|\boldsymbol{X}]-\beta_1^*-\beta_0^*\right)\cdot\frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot p\cdot\begin{bmatrix}1\\1\end{bmatrix}\middle|\boldsymbol{X}\right] \\
&+\mathbb{E}\left[\left(\mathbb{E}[Y^*(0)|\boldsymbol{X}]-\beta_0^*\right)\cdot\frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}\cdot q\cdot\begin{bmatrix}1\\0\end{bmatrix}\middle|\boldsymbol{X}\right] \\
=&\begin{bmatrix}\left(\mathbb{E}[Y^*(1)|\boldsymbol{X}]-\beta_1^*-\beta_0^*\right)\cdot p+\left(\mathbb{E}[Y^*(0)|\boldsymbol{X}]-\beta_1^*\right)\cdot q \\ \left(\mathbb{E}[Y^*(1)|\boldsymbol{X}]-\beta_1^*-\beta_0^*\right)\cdot p\end{bmatrix}.
\end{aligned} \tag{9}
$$

Therefore, with (7), (8), and (9) we can obtain that

$$
\begin{aligned}
&\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\left\{Y-\mathbb{E}[Y|\boldsymbol{X},T]\right\}+\mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|\boldsymbol{X}\right] \\
=&\begin{pmatrix}p\cdot\phi_1(T,\boldsymbol{X},Y;\boldsymbol{\beta}^*)+q\cdot\phi_2(T,\boldsymbol{X},Y;\boldsymbol{\beta}^*) \\ p\cdot\phi_1(T,\boldsymbol{X},Y;\boldsymbol{\beta}^*)\end{pmatrix},
\end{aligned}
$$

and the efficient influence functions of β_0^* and β_1^* are given by

$$
\begin{bmatrix}\frac{1}{q} & -\frac{1}{q} \\ -\frac{1}{q} & \frac{1}{pq}\end{bmatrix}\cdot\begin{pmatrix}p\cdot\phi_1(T,\boldsymbol{X},Y;\boldsymbol{\beta})+q\cdot\phi_2(T,\boldsymbol{X},Y;\boldsymbol{\beta}^*) \\ p\cdot\phi_1(T,\boldsymbol{X},Y;\boldsymbol{\beta}^*)\end{pmatrix}=\begin{pmatrix}\phi_2(T,\boldsymbol{X},Y;\boldsymbol{\beta}^*) \\ \phi_1(T,\boldsymbol{X},Y;\boldsymbol{\beta}^*)-\phi_2(T,\boldsymbol{X},Y;\boldsymbol{\beta}^*)\ .\end{pmatrix}.
$$

■

2.3 Particular Case II: Multiple Average Treatment Effects

In this section, we show that when $T \in \{0, 1, ..., J\}$, $J \in \mathbb{N}$, $g(t; \boldsymbol{\beta}) = \sum_{j=0}^{J} \beta_j \cdot I(t = j)$ and $L(v) = v^2$, our general efficiency bound derived in Theorem 1 reduces to the efficiency bound of multi-level treatment effects given in Cattaneo (2010). In accordance with our proposed identification condition, $\{\beta_j^*\}_{j=0}^{J}$ are identified by minimizing the following loss function

$$\sum_{j=0}^{J} \mathbb{E}\left[(Y^*(j) - \beta_j)^2\right] \cdot \mathbb{P}(T = j).$$

The solutions are $\beta_j^* = \mathbb{E}[Y^*(j)]$ for $j \in \{0, ..., J\}$.

Corollary 2.2 *Suppose $T \in \{0, 1, ..., J\}$, $J \in \mathbb{N}$, $g(t; \boldsymbol{\beta}) = \sum_{j=0}^{J} \beta_j \cdot I(t = j)$, $L(v) = v^2$, and the conditions in Theorem 1 hold, the efficient influence functions of $\{\beta_j^*\}_{j=0}^{J}$ given by Theorem 1 reduce to*

$$S_{eff}(T, \boldsymbol{X}, Y; \beta_j^*) = \frac{I(T = j)}{\mathbb{P}(T = j|\boldsymbol{X})} \cdot \{Y^*(j) - \mathbb{E}[Y^*(j)|\boldsymbol{X}]\} + \mathbb{E}[Y^*(j)|X] - \beta_j^*, \ j \in \{0, ..., J\},$$

and they are the same as the efficient influence functions given in Cattaneo (2010).

Proof. Using our notation, we have

$$\boldsymbol{\beta}^* = (\beta_0^*, ..., \beta_J^*)^\top, \ g(t; \boldsymbol{\beta}^*) = \sum_{j=0}^{J} \beta_j \cdot I(t = j), \ m(t; \boldsymbol{\beta}^*) = \begin{bmatrix} I(t = 0) \\ I(t = 1) \\ \vdots \\ I(t = J) \end{bmatrix}, \ H_0 = \mathbb{E}\left[m(T; \boldsymbol{\beta}^*) m(T; \boldsymbol{\beta}^*)^\top\right].$$

Then

$$\begin{aligned}
\varepsilon(T, \boldsymbol{X}; \boldsymbol{\beta}^*) =& \mathbb{E}[Y|T, X] - g(T; \boldsymbol{\beta}^*) \\
=& \sum_{j=0}^{J} \mathbb{E}[Y^*(j)|X] \cdot I(t = j) - \sum_{j=0}^{J} \beta_j^* \cdot I(T = j) \\
=& \sum_{j=0}^{J} \left(\mathbb{E}[Y^*(j)|X] - \beta_j^*\right) \cdot I(T = j)
\end{aligned}$$

and

$$\pi_0(T, X) = \sum_{j=0}^{J} \frac{I(T=j)}{\mathbb{P}(T=j|\boldsymbol{X})} \cdot p_j, \text{ where } p_j = \mathbb{P}(T=j).$$

Then we have

$$H_0^{-1} = \mathbb{E}\left[m(T;\boldsymbol{\beta}^*)m(T;\boldsymbol{\beta}^*)^\top\right]^{-1} = \begin{bmatrix} p_0^{-1} & & & \\ & p_1^{-1} & & \\ & & \cdots & \\ & & & p_J^{-1} \end{bmatrix},$$

and

$$\begin{aligned}
&\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\left\{Y - \mathbb{E}[Y|\boldsymbol{X},T]\right\} \\
=&\left\{\sum_{j=0}^{J} \frac{I(T=j)}{\mathbb{P}(T=j|\boldsymbol{X})} \cdot p_j\right\} \cdot \begin{bmatrix} I(T=0) \\ I(T=1) \\ \vdots \\ I(T=J) \end{bmatrix} \cdot \left\{Y - \sum_{j=0}^{J} I(T=j) \cdot \mathbb{E}[Y^*(j)|\boldsymbol{X}]\right\} \\
=&\begin{bmatrix} I(T=0) \\ I(T=1) \\ \vdots \\ I(T=J) \end{bmatrix} \left\{\sum_{j=0}^{J} \frac{I(T=j)}{\mathbb{P}(T=j|\boldsymbol{X})} \cdot p_j \cdot Y^*(j) - \sum_{j=0}^{J} \frac{I(T=j)}{\mathbb{P}(T=j|\boldsymbol{X})} \cdot p_j \cdot \mathbb{E}[Y^*(j)|\boldsymbol{X}]\right\} \\
=&\begin{bmatrix} \frac{I(T=0)}{\mathbb{P}(T=0|\boldsymbol{X})} \cdot p_0 \cdot \{Y^*(0) - \mathbb{E}[Y^*(0)|\boldsymbol{X}]\} \\ \frac{I(T=1)}{\mathbb{P}(T=1|\boldsymbol{X})} \cdot p_1 \cdot \{Y^*(1) - \mathbb{E}[Y^*(1)|\boldsymbol{X}]\} \\ \vdots \\ \frac{I(T=J)}{\mathbb{P}(T=J|\boldsymbol{X})} \cdot p_J \cdot \{Y^*(j) - \mathbb{E}[Y^*(j)|\boldsymbol{X}]\} \end{bmatrix} \qquad (10)
\end{aligned}$$

and

$$\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)$$

$$= \left\{ \sum_{j=0}^{J} \left(\mathbb{E}[Y^*(j)|X] - \beta_j^* \right) \cdot I(T=j) \right\} \left\{ \sum_{j=0}^{J} \frac{I(T=j)}{\mathbb{P}(T=j|\boldsymbol{X})} \cdot p_j \right\} \begin{bmatrix} I(T=0) \\ I(T=1) \\ \vdots \\ I(T=J) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{I(T=0)}{\mathbb{P}(T=0|\boldsymbol{X})} \cdot p_0 \cdot \{\mathbb{E}[Y^*(0)|X] - \beta_0^*\} \\ \frac{I(T=1)}{\mathbb{P}(T=1|\boldsymbol{X})} \cdot p_1 \cdot \{\mathbb{E}[Y^*(1)|X] - \beta_1^*\} \\ \vdots \\ \frac{I(T=J)}{\mathbb{P}(T=J|\boldsymbol{X})} \cdot p_J \cdot \{\mathbb{E}[Y^*(j)|X] - \beta_J^*\} \end{bmatrix}$$

and

$$\mathbb{E}\left[\varepsilon(T, \boldsymbol{X}; \boldsymbol{\beta}^*)\pi_0(T, \boldsymbol{X}) m(T; \boldsymbol{\beta}^*)|\boldsymbol{X}\right] = \begin{bmatrix} p_0 \cdot \{\mathbb{E}[Y^*(0)|X] - \beta_0^*\} \\ p_1 \cdot \{\mathbb{E}[Y^*(1)|X] - \beta_1^*\} \\ \vdots \\ p_J \cdot \{\mathbb{E}[Y^*(j)|X] - \beta_J^*\} \end{bmatrix}. \quad (11)$$

From Theorem 1, the efficient influence function of $\boldsymbol{\beta}^* = (\beta_0^*, ..., \beta_J^*)$ is given by

$$H_0^{-1} \left\{ \pi_0(T, \boldsymbol{X}) m(T; \boldsymbol{\beta}^*) \left\{ Y - \mathbb{E}[Y|\boldsymbol{X}, T] \right\} + \mathbb{E}\left[\varepsilon(T, \boldsymbol{X}; \boldsymbol{\beta}_0)\pi_0(T, \boldsymbol{X}) m(T; \boldsymbol{\beta}^*)|\boldsymbol{X}\right] \right\}$$

$$= \begin{bmatrix} \frac{I(T=0)}{\mathbb{P}(T=0|\boldsymbol{X})} \cdot \{Y^*(0) - \mathbb{E}[Y^*(0)|\boldsymbol{X}]\} + \mathbb{E}[Y^*(0)|X] - \beta_0^* \\ \frac{I(T=1)}{\mathbb{P}(T=1|\boldsymbol{X})} \cdot \{Y^*(1) - \mathbb{E}[Y^*(1)|\boldsymbol{X}]\} + \mathbb{E}[Y^*(1)|X] - \beta_1^* \\ \vdots \\ \frac{I(T=J)}{\mathbb{P}(T=J|\boldsymbol{X})} \cdot \{Y^*(j) - \mathbb{E}[Y^*(j)|\boldsymbol{X}]\} + \mathbb{E}[Y^*(j)|X] - \beta_J^* \end{bmatrix},$$

which is the same as the efficient influence function developed in Corollary 1 of Cattaneo (2010). ■

2.4 Particular Case III: Binary Quantile Treatment Effects

In this section, we show that when $T \in \{0, 1\}$ is a binary treatment variable, $L(v) = v(\tau - I(v \leq 0))$ is the check function with $\tau \in (0, 1)$, and $g(t; \boldsymbol{\beta}^*) = \beta_0^* \cdot (1-t) + \beta_1^* \cdot t$, where $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1^*)$, our general efficiency bound derived in Theorem 1 reduces to the efficiency bound of quantile treatment effects given in Firpo (2007). In accordance with our identification condition, β_0^*

and β_1^* are identified by minimizing the following loss function

$$\sum_{j\in\{0,1\}} \mathbb{P}(T=j)\cdot\mathbb{E}\left[(Y^*(j)-\beta_j)\left\{\tau - I(Y^*(j)\le\beta_j)\right\}\right].$$

The solutions are $\beta_0^* = \inf\{q:\mathbb{P}(Y^*(0)\le q)\ge\tau\}$ and $\beta_1^* = \inf\{q:\mathbb{P}(Y^*(1)\le q)\ge\tau\}$, which are the τ^{th} quantiles of potential outcomes.

Corollary 2.3 *Let $T\in\{0,1\}$, $f_{Y^*(1)}$ and $f_{Y^*(0)}$ be the probability densities of the potential outcomes $Y^*(1)$ and $Y^*(0)$ respectively, $g(t;\boldsymbol{\beta}^*) = \beta_0^*\cdot(1-t)+\beta_1^*\cdot t$, $L(v) = v(\tau - I(v\le 0))$, and the conditions in Theorem 1 hold, then the efficient influence function of $\boldsymbol{\beta}^*$ given by Theorem 1 reduces to*

$$S_{eff}(Y,T,\boldsymbol{X};\boldsymbol{\beta}^*) = \begin{bmatrix} \frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}\cdot\left\{\frac{\tau-I(Y^*(0)\le\beta_0^*)}{f_{Y^*(0)}(\beta_0^*)}\right\} - \left(\frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}-1\right)\cdot\mathbb{E}\left[\frac{\tau-I(Y^*(0)\le\beta_0^*)}{f_{Y^*(0)}(\beta_0^*)}|\boldsymbol{X}\right] \\ \frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot\left\{\frac{\tau-I(Y^*(1)\le\beta_1^*)}{f_{Y^*(1)}(\beta_1^*)}\right\} - \left(\frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}-1\right)\cdot\mathbb{E}\left[\frac{\tau-I(Y^*(1)\le\beta_1^*)}{f_{Y^*(1)}(\beta_1^*)}|\boldsymbol{X}\right]\end{bmatrix},$$

which is the same as the efficient influence function given in Firpo (2007).

Proof. Using our notation, we have

$$\begin{aligned}
&\boldsymbol{\beta}^* = (\beta_0^*,\beta_1^*)^\top,\ g(t;\boldsymbol{\beta}^*) = \beta_0^*\cdot(1-t)+\beta_1^*\cdot t,\ m(t;\boldsymbol{\beta}^*) = \begin{bmatrix}1-t\\ t\end{bmatrix},\\
&L(v) = v(\tau - I(v\le 0)),\ L'(v) = \tau - I(v\le 0)\text{ a.s.},\\
&\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*) = T\cdot\mathbb{E}[\tau - I(Y^*(1)\le\beta_1^*)|\boldsymbol{X}] + (1-T)\cdot\mathbb{E}[\tau - I(Y^*(0)\le\beta_0^*)|\boldsymbol{X}],\\
&\pi_0(T,\boldsymbol{X}) = \frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot p + \frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}\cdot q\ ,\ p=\mathbb{P}(T=1),\ q=\mathbb{P}(T=0).
\end{aligned}$$

Direct computation yields

$$\begin{aligned}
&\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)L'(Y-g(T;\boldsymbol{\beta}^*))\\
=&\left\{\frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot p + \frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}\cdot q\right\}\cdot\begin{bmatrix}1-T\\ T\end{bmatrix}\cdot\left\{\tau - I(Y\le\beta_0^*\cdot(1-T)+\beta_1^*\cdot T)\right\}\\
=&\begin{bmatrix}\frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}\cdot q\cdot\{\tau - I(Y^*(0)\le\beta_0^*)\}\\ \frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot p\cdot\{\tau - I(Y^*(1)\le\beta_1^*)\}\end{bmatrix}
\end{aligned}$$

and

$$\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*) = \begin{bmatrix} \frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}\cdot q\cdot \mathbb{E}\left[\tau - I(Y^*(0)\le \beta_0^*)|\boldsymbol{X}\right] \\ \frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot p\cdot \mathbb{E}\left[\tau - I(Y^*(1)\le \beta_1^*)|\boldsymbol{X}\right] \end{bmatrix}$$

and

$$\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)|\boldsymbol{X}\right] = \begin{bmatrix} q\cdot \mathbb{E}\left[\tau - I(Y^*(0)\le \beta_0^*)|\boldsymbol{X}\right] \\ p\cdot \mathbb{E}\left[\tau - I(Y^*(1)\le \beta_1^*)|\boldsymbol{X}\right] \end{bmatrix}$$

and

$$H_0 = \nabla_{\boldsymbol{\beta}}\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)L'(Y-g(T;\boldsymbol{\beta}^*))\right] = \begin{bmatrix} -q\cdot f_{Y^*(0)}(\beta_0^*) & 0 \\ 0 & -p\cdot f_{Y^*(1)}(\beta_1^*) \end{bmatrix}.$$

Therefore, by Theorem 1, the efficient influence function of $\boldsymbol{\beta}^*$ is

$$\begin{aligned}
&S_{eff}(Y,T,\boldsymbol{X};\boldsymbol{\beta}^*) \\
=&H_0^{-1}\cdot\Big\{\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)L'(Y-g(T;\boldsymbol{\beta}^*)) - \pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*) \\
&\qquad + \mathbb{E}\left[\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)|\boldsymbol{X}\right]\Big\} \\
=&\begin{bmatrix} q^{-1}\cdot\frac{1}{f_{Y^*(0)}(\beta_0^*)} & 0 \\ 0 & p^{-1}\cdot\frac{1}{f_{Y^*(1)}(\beta_1^*)} \end{bmatrix} \\
&\times\begin{bmatrix} \frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}\cdot q\cdot\{\tau - I(Y^*(0)\le\beta_0^*)\} - q\cdot\left(\frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}-1\right)\cdot\mathbb{E}\left[\tau - I(Y^*(0)\le \beta_0^*)|\boldsymbol{X}\right] \\ \frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot p\cdot\{\tau - I(Y^*(1)\le\beta_1^*)\} - p\cdot\left(\frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}-1\right)\cdot\mathbb{E}\left[\tau - I(Y^*(1)\le \beta_1^*)|\boldsymbol{X}\right] \end{bmatrix} \\
=&\begin{bmatrix} \frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}\cdot\left\{\frac{\tau-I(Y^*(0)\le\beta_0^*)}{f_{Y^*(0)}(\beta_0^*)}\right\} - \left(\frac{1-T}{\mathbb{P}(T=0|\boldsymbol{X})}-1\right)\cdot\mathbb{E}\left[\frac{\tau-I(Y^*(0)\le\beta_0^*)}{f_{Y^*(0)}(\beta_0^*)}\middle|\boldsymbol{X}\right] \\ \frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}\cdot\left\{\frac{\tau-I(Y^*(1)\le\beta_1^*)}{f_{Y^*(1)}(\beta_1^*)}\right\} - \left(\frac{T}{\mathbb{P}(T=1|\boldsymbol{X})}-1\right)\cdot\mathbb{E}\left[\frac{\tau-I(Y^*(1)\le\beta_1^*)}{f_{Y^*(1)}(\beta_1^*)}\middle|\boldsymbol{X}\right] \end{bmatrix},
\end{aligned}$$

which coincides with efficiency bound derived in Firpo (2007). ■

3 Convergence Rate of Estimated Stabilized Weights

In this section, we establish the convergence rate of estimated stabilized weights $\hat{\pi}_K(T,\boldsymbol{X})$. Let $G^*_{K_1\times K_2}$, $\Lambda^*_{K_1\times K_2}$ and $\pi^*_K(t,\boldsymbol{x})$ be the theoretical counterparts of $\hat{G}_{K_1\times K_2}$, $\hat{\Lambda}_{K_1\times K_2}$ and $\hat{\pi}_K(t,\boldsymbol{x})$

respectively:

$$\begin{aligned}
G^*_{K_1\times K_2}(\Lambda) :=&\mathbb{E}[\hat{G}_{K_1\times K_2}(\Lambda)] = \mathbb{E}\left[\rho\left(u_{K_1}(T)^\top \Lambda v_{K_2}(\boldsymbol{X})\right)\right] - \mathbb{E}[u_{K_1}(T)^\top]\cdot\Lambda\cdot\mathbb{E}[v_{K_2}(\boldsymbol{X})],\\
\Lambda^*_{K_1\times K_2} :=& \arg\max G^*_{K_1\times K_2}(\Lambda),\\
\pi^*_K(t,\boldsymbol{x}) :=&\rho'\left(u_{K_1}(t)^\top \Lambda^*_{K_1\times K_2} v_{K_2}(\boldsymbol{x})\right).
\end{aligned}$$

Because of Assumption 1.4, without loss of generality, we can assume the sieve bases $u_{K_1}(T)$ and $v_{K_2}(\boldsymbol{X})$ are orthonormalized, i.e.,

$$\mathbb{E}\left[u_{K_1}(T)u_{K_1}^\top(T)\right] = I_{K_1\times K_1}, \quad \mathbb{E}\left[v_{K_2}(\boldsymbol{X})v_{K_2}^\top(\boldsymbol{X})\right] = I_{K_2\times K_2}. \tag{12}$$

Let

$$\zeta_1(K_1) := \sup_{t\in\mathcal{T}}\|u_{K_1}(t)\|\ ,\ \zeta_2(K_2) := \sup_{\boldsymbol{x}\in\mathcal{X}}\|v_{K_2}(\boldsymbol{x})\|\ ,\ K = K_1\cdot K_2\ ,\ \zeta(K) = \zeta_1(K_1)\zeta_2(K_2).$$

We also recall the following property satisfied by $\pi_0(T,\boldsymbol{X})$: for any integrable functions $u(t)$ and $v(\boldsymbol{X})$,

$$\mathbb{E}\left[\pi_0(T,\boldsymbol{X})u(T)v(\boldsymbol{X})\right] = \mathbb{E}[u(T)]\cdot\mathbb{E}[v(\boldsymbol{X})]. \tag{13}$$

3.1 Lemma 3.1

The first lemma states that $\pi^*_K(t,\boldsymbol{x})$ is arbitrarily close to the true stabilized weights $\pi_0(t,\boldsymbol{x})$.

Lemma 3.1 *Under Assumption 1.2-1.4, we have*

$$\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}|\pi_0(t,\boldsymbol{x}) - \pi^*_K(t,\boldsymbol{x})| = O(\zeta(K)K^{-\alpha}),$$

and

$$\mathbb{E}\left[|\pi_0(T,\boldsymbol{X}) - \pi^*_K(T,\boldsymbol{X})|^2\right] = O\left(K^{-2\alpha}\right),$$

and

$$\frac{1}{N}\sum_{i=1}^N|\pi_0(T_i,\boldsymbol{X}_i) - \pi^*_K(T_i,\boldsymbol{X}_i)|^2 = O_p\left(K^{-2\alpha}\right).$$

Proof. By Assumption 1.2, $\pi_0(t,\boldsymbol{x}) \in [\eta_1, \eta_2]$, $\forall (t,\boldsymbol{x}) \in \mathcal{T}\times\mathcal{X}$ and $(\rho')^{-1}$ is strictly decreasing. Define

$$\overline{\gamma} := \sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} (\rho')^{-1}\left(\pi_0(t,\boldsymbol{x})\right) \le (\rho')^{-1}(\eta_1) \text{ and } \underline{\gamma} := \inf_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} (\rho')^{-1}\left(\pi_0(t,\boldsymbol{x})\right) \ge (\rho')^{-1}(\eta_2),$$

which are two finite constants. By Assumptions 1.3, there exist a constant $C > 0$ and a $K_1 \times K_2$ matrix $\Lambda_{K_1\times K_2} \in \mathbb{R}^{K_1\times K_2}$ such that

$$\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} \left|(\rho')^{-1}\left(\pi_0(t,\boldsymbol{x})\right) - u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x})\right| < CK^{-\alpha},$$

which implies

$$\begin{aligned} u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x}) &\in \left((\rho')^{-1}\left(\pi_0(t,\boldsymbol{x})\right) - CK^{-\alpha}, (\rho')^{-1}\left(\pi_0(t,\boldsymbol{x})\right) + CK^{-\alpha}\right) \\ &\subset \left[\underline{\gamma} - CK^{-\alpha}, \overline{\gamma} + CK^{-\alpha}\right],\ \forall (t,\boldsymbol{x}) \in \mathcal{T}\times\mathcal{X}, \end{aligned} \tag{14}$$

and

$$\begin{aligned} &\rho'\left(u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x}) + CK^{-\alpha}\right) - \rho'(u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{X})) \\ <&\pi_0(t,\boldsymbol{x}) - \rho'\left(u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x})\right) \\ <&\rho'\left(u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x}) - CK^{-\alpha}\right) - \rho'(u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x}))\ , \forall (t,\boldsymbol{x}) \in \mathcal{T}\times\mathcal{X}. \end{aligned}$$

Let $\Gamma_1 := [\underline{\gamma} - 1, \overline{\gamma} + 1]$, by Mean Value Theorem, for large enough K, there exist

$$\begin{aligned} \xi_1(t,\boldsymbol{x}) &\in \left(u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x}), u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x}) + CK^{-\alpha}\right) \\ &\subset \left[\underline{\gamma} - CK^{-\alpha}, \overline{\gamma} + 2CK^{-\alpha}\right] \subset \Gamma_1\ , \\ \xi_2(t,\boldsymbol{x}) &\in \left(u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x}) - CK^{-\alpha}, u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x})\right) \\ &\subset \left[\underline{\gamma} - 2CK^{-\alpha}, \overline{\gamma} + CK^{-\alpha}\right] \subset \Gamma_1, \end{aligned}$$

such that

$$\rho'\left(u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x}) + CK^{-\alpha}\right) - \rho'\left(u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x})\right) = \rho''(\xi_1(t,x))CK^{-\alpha} \ge -a_1 CK^{-\alpha}$$

and

$$\rho'\left(u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x}) - CK^{-\alpha}\right) - \rho'\left(u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x})\right) = -\rho''(\xi_2(t,\boldsymbol{x}))CK^{-\alpha} \le a_2 CK^{-\alpha},$$

where $-a_1 := \inf_{\gamma\in\Gamma_1} \rho''(\gamma)$ and $a_2 := \sup_{\gamma\in\Gamma_1}(-\rho''(\gamma))$. Let $a := \max\{a_1, a_2\}$, we have

$$\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} \left|\pi_0(t,\boldsymbol{x}) - \rho'\left(u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x})\right)\right| < aCK^{-\alpha}. \tag{15}$$

For some fixed $C_2 > 0$ (to be chosen later), define

$$\Upsilon_{K_1\times K_2} := \left\{\Lambda \in \mathbb{R}^{K_1\times K_2} : \|\Lambda - \Lambda_{K_1\times K_2}\| \le C_2 K^{-\alpha}\right\}.$$

For sufficiently large K_1 and K_2, we have that $\forall\Lambda \in \Upsilon_{K_1\times K_2}$, $\forall(t,\boldsymbol{x}) \in \mathcal{T}\times\mathcal{X}$,

$$\begin{aligned}&\left|u_{K_1}(t)^\top \Lambda v_{K_2}(\boldsymbol{x}) - u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x})\right| \\ \le&\|\Lambda - \Lambda_{K_1\times K_2}\| \cdot \sup_{\boldsymbol{x}\in\mathcal{X}} \|v_{K_2}(\boldsymbol{x})\| \cdot \sup_{t\in\mathcal{T}} \|u_{K_1}(t)\| \le C_2 K^{-\alpha}\zeta_1(K_1)\zeta_2(K_2).\end{aligned}$$

Then in light of (14) and Assumption 1.4, for large enough K_1 and K_2, $\forall\Lambda \in \Upsilon_{K_1\times K_2}$ and $\forall(t,\boldsymbol{x}) \in \mathcal{T}\times\mathcal{X}$, we can deduce that

$$\begin{aligned}u_{K_1}(t)^\top \Lambda v_{K_2}(\boldsymbol{x}) \in &\Big(u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x}) - C_2 K^{-\alpha}\zeta_1(K_1)\zeta_2(K_2), \\ &\qquad u_{K_1}(t)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{x}) + C_2 K^{-\alpha}\zeta_1(K_1)\zeta_2(K_2)\Big) \\ &\subset \Big[\underline{\gamma} - CK^{-\alpha} - C_2 K^{-\alpha}\zeta_1(K_1)\zeta_2(K_2), \\ &\qquad \overline{\gamma} + CK^{-\alpha} + C_2 K^{-\alpha}\zeta_1(K_1)\zeta_2(K_2)\Big] \subset \Gamma_1.\end{aligned} \tag{16}$$

By definition

$$G^*_{K_1\times K_2}(\Lambda) = \mathbb{E}\left[\rho\left(u_{K_1}(T)^\top \Lambda v_{K_2}(\boldsymbol{X})\right)\right] - \mathbb{E}[u_{K_1}(T)]^\top \Lambda \mathbb{E}[v_{K_2}(\boldsymbol{X})],$$

is a strictly concave function of Λ. By (13), the formula $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for matrices A and B, the facts $\mathbb{E}\left[v_{K_2}(\boldsymbol{X})v_{K_2}(\boldsymbol{X})^\top\right] = I_{K_2\times K_2}$ and $\mathbb{E}\left[u_{K_1}(T)u_{K_1}(T)^\top\right] = I_{K_1\times K_1}$, we can deduce that

$$\begin{aligned}&\|\nabla G^*_{K_1\times K_2}(\Lambda_{K_1\times K_2})\|^2 \\ =&\left\|\mathbb{E}\left[\rho'\left(u_{K_1}(T)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{X})\right) u_{K_1}(T) v_{K_2}(\boldsymbol{X})^\top\right] - \mathbb{E}[u_{K_1}(T)]\mathbb{E}[v_{K_2}(\boldsymbol{X})]^\top\right\|^2 \\ =&\left\|\mathbb{E}\left[\rho'\left(u_{K_1}(T)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{X})\right) u_{K_1}(T) v_{K_2}(\boldsymbol{X})^\top\right] - \mathbb{E}[\pi_0(T,\boldsymbol{X}) u_{K_1}(T) v_{K_2}(\boldsymbol{X})]^\top\right\|^2 \quad \text{(by (13))} \\ =&\left\|\mathbb{E}\left[\sqrt{\pi_0(T,\boldsymbol{X})}\frac{\left\{\rho'\left(u_{K_1}(T)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{X})\right) - \pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}} u_{K_1}(T) v_{K_2}(\boldsymbol{X})^\top\right]\right\|^2 \\ =&\mathrm{tr}\Bigg\{\mathbb{E}\left[\sqrt{\pi_0(T,\boldsymbol{X})}\frac{\left\{\rho'\left(u_{K_1}(T)^\top \Lambda_{K_1\times K_2} v_{K_2}(\boldsymbol{X})\right) - \pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}} u_{K_1}(T) v_{K_2}(\boldsymbol{X})^\top\right]\end{aligned}$$

$$
\begin{aligned}
&\times \mathbb{E}\left[\sqrt{\pi_0(T,\boldsymbol{X})}\frac{\left\{\rho'\left(u_{K_1}(T)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)-\pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}}v_{K_2}(\boldsymbol{X})u_{K_1}(T)^\top\right]\Bigg\}\\
=&\operatorname{tr}\Bigg\{\mathbb{E}\left[\sqrt{\pi_0(T,\boldsymbol{X})}\frac{\left\{\rho'\left(u_{K_1}(T)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)-\pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}}u_{K_1}(T)v_{K_2}(\boldsymbol{X})^\top\right]\cdot\mathbb{E}\left[u_{K_2}(\boldsymbol{X})u_{K_2}(\boldsymbol{X})^\top\right]\\
&\times \mathbb{E}\left[\sqrt{\pi_0(T,\boldsymbol{X})}\frac{\left\{\rho'\left(u_{K_1}(T)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)-\pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}}v_{K_2}(\boldsymbol{X})u_{K_1}(T)^\top\right]\cdot\mathbb{E}\left[u_{K_1}(T)u_{K_1}(T)^\top\right]\Bigg\}\\
=&\mathbb{E}\Bigg[\operatorname{tr}\Bigg\{u_{K_1}(T)^\top\cdot\mathbb{E}\left[\sqrt{\pi_0(T,\boldsymbol{X})}\frac{\left\{\rho'\left(u_{K_1}(T)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)-\pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}}u_{K_1}(T)v_{K_2}(\boldsymbol{X})^\top\right]\cdot\mathbb{E}\left[u_{K_2}(\boldsymbol{X})u_{K_2}(\boldsymbol{X})^\top\right]\\
&\times \mathbb{E}\left[\sqrt{\pi_0(T,\boldsymbol{X})}\frac{\left\{\rho'\left(u_{K_1}(T)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)-\pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}}v_{K_2}(\boldsymbol{X})u_{K_1}(T)^\top\right]\cdot u_{K_1}(T)\Bigg\}\Bigg]\\
=&\mathbb{E}\Bigg[\pi_0(T,\boldsymbol{X})\cdot u_{K_1}(T)^\top\cdot\mathbb{E}\left[\sqrt{\pi_0(T,\boldsymbol{X})}\frac{\left\{\rho'\left(u_{K_1}(T)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)-\pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}}u_{K_1}(T)v_{K_2}(\boldsymbol{X})^\top\right]\cdot u_{K_2}(\boldsymbol{X})\\
&\times\cdot u_{K_2}(\boldsymbol{X})^\top\mathbb{E}\left[\sqrt{\pi_0(T,\boldsymbol{X})}\frac{\left\{\rho'\left(u_{K_1}(T)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)-\pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}}v_{K_2}(\boldsymbol{X})u_{K_1}(T)^\top\right]\cdot u_{K_1}(T)\Bigg]\quad\text{(by (13))}\\
=&\mathbb{E}\left[\left|\pi_0(T,\boldsymbol{X})^{\frac14}u_{K_1}(T)\mathbb{E}\left[\sqrt{\pi_0(T,\boldsymbol{X})}\frac{\left\{\rho'\left(u_{K_1}(T)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)-\pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}}u_{K_1}(T)v_{K_2}(\boldsymbol{X})^\top\right]\pi_0(T,\boldsymbol{X})^{\frac14}v_{K_2}(\boldsymbol{X})\right|^2\right].
\end{aligned}
\tag{17}
$$

Note that the term in the last expression

$$
\pi_0(T,\boldsymbol{X})^{\frac14}u_{K_1}(T)\cdot\mathbb{E}\left[\sqrt{\pi_0(T,\boldsymbol{X})}\frac{\left\{\rho'\left(u_{K_1}(T)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)-\pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}}u_{K_1}(T)v_{K_2}(\boldsymbol{X})^\top\right]\pi_0(T,\boldsymbol{X})^{\frac14}v_{K_2}(\boldsymbol{X})
$$

is the $L^2(dF_{T,X})$-projection of $\frac{\left\{\rho'\left(u_{K_1}(T)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)-\pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}}$ on the space spanned by $\{\pi_0(T,\boldsymbol{X})^{\frac14}u_{K_1}(T),\ \pi_0(T,\boldsymbol{X})^{\frac14}v_{K_2}(\boldsymbol{X})\}$, which implies that

$$
\begin{aligned}
&\mathbb{E}\left[\left|\pi_0(T,\boldsymbol{X})^{\frac14}u_{K_1}(T)\mathbb{E}\left[\sqrt{\pi_0(T,\boldsymbol{X})}\frac{\left\{\rho'\left(u_{K_1}(T)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)-\pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}}u_{K_1}(T)v_{K_2}(\boldsymbol{X})^\top\right]\pi_0(T,\boldsymbol{X})^{\frac14}v_{K_2}(\boldsymbol{X})\right|^2\right]\\
\le&\mathbb{E}\left[\left|\frac{\left\{\rho'\left(u_{K_1}(T)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)-\pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}}\right|^2\right].
\end{aligned}
\tag{18}
$$

Now, with (17), (18), we can obtain that

$$
\begin{aligned}
&\|\nabla G^*_{K_1\times K_2}(\Lambda_{K_1\times K_2})\|\\
\le&\mathbb{E}\left[\left|\frac{\left\{\rho'\left(u_{K_1}(T)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)-\pi_0(T,\boldsymbol{X})\right\}}{\sqrt{\pi_0(T,\boldsymbol{X})}}\right|^2\right]^{\frac12}\\
\le&\frac{1}{\sqrt{\eta_1}}\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}\left|\rho'\left(u_{K_1}(t)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)-\pi_0(t,\boldsymbol{x})\right|\qquad\text{(by Assumption 1.2)}
\end{aligned}
$$

$$\leq \frac{aC}{\sqrt{\eta_1}} \cdot K^{-\alpha} \qquad \text{(by (15))}. \tag{19}$$

Note that for any $\Lambda \in \partial\Upsilon_{K_1\times K_2}$, i.e. $\|\Lambda - \Lambda_{K_1\times K_2}\| = C_2K^{-\alpha}$, by Mean Value Theorem and the fact $\rho''(y) = -\rho'(y)$, we can deduce that

$$\begin{aligned}
& G^*_{K_1\times K_2}(\Lambda) - G^*_{K_1\times K_2}(\Lambda_{K_1\times K_2}) \\
&= \sum_{j=1}^{K_2}(\lambda_j - \lambda_j^K)^\top \frac{\partial}{\partial\lambda_i} G^*_{K_1\times K_2}(\lambda_1^K, \ldots, \lambda_{K_2}^K) \\
&\quad + \sum_{l=1}^{K_2}\sum_{j=1}^{K_2}\frac{1}{2}(\lambda_j - \lambda_j^K)^\top \frac{\partial^2}{\partial\lambda_i\partial\lambda_l} G^*_{K_1\times K_2}(\bar\lambda_1^K, \ldots, \bar\lambda_{K_2}^K)(\lambda_l - \lambda_l^K) \\
&\leq \|\Lambda - \Lambda_{K_1\times K_2}\| \, \|\nabla G^*_{K_1\times K_2}(\Lambda_{K_1\times K_2})\| \\
&+ \frac{1}{2}\sum_{l=1}^{K_2}\sum_{j=1}^{K_2}(\lambda_j - \lambda_j^K)^\top \mathbb{E}\left[\rho''\left(u_{K_1}^\top(T)\bar\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)u_{K_1}(T)u_{K_1}(T)^\top v_{K_2,j}(\boldsymbol{X})v_{K_2,l}(\boldsymbol{X})\right](\lambda_l - \lambda_l^K) \\
&= \|\Lambda - \Lambda_{K_1\times K_2}\| \, \|\nabla G^*_{K_1\times K_2}(\Lambda_{K_1\times K_2})\| \\
&\quad - \frac{1}{2}\sum_{l=1}^{K_2}\sum_{j=1}^{K_2}(\lambda_j - \lambda_j^K)^\top \mathbb{E}\left[\frac{\rho'\left(u_{K_1}^\top(T)\bar\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)}{\pi_0(T,\boldsymbol{X})}\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^\top v_{K_2,j}(\boldsymbol{X})v_{K_2,l}(\boldsymbol{X})\right](\lambda_l - \lambda_l^K) \\
&\leq \|\Lambda - \Lambda_{K_1\times K_2}\| \, \|\nabla G^*_{K_1\times K_2}(\Lambda_{K_1\times K_2})\| \\
&\quad - \frac{a_3}{2\eta_2}\sum_{l=1}^{K_2}\sum_{j=1}^{K_2}(\lambda_j - \lambda_j^K)^\top \mathbb{E}\left[\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^\top v_{K_2,j}(\boldsymbol{X})v_{K_2,l}(\boldsymbol{X})\right](\lambda_l - \lambda_l^K) \quad \left(\text{by } a_3 = \inf_{y\in\Gamma_1}\{\rho'(y)\}\right) \\
&= \|\Lambda - \Lambda_{K_1\times K_2}\| \, \|\nabla G^*_{K_1\times K_2}(\Lambda_{K_1\times K_2})\| \\
&\quad - \frac{a_3}{2\eta_2}\sum_{l=1}^{K_2}\sum_{j=1}^{K_2}(\lambda_j - \lambda_j^K)^\top \mathbb{E}\left[u_{K_1}(T)u_{K_1}(T)^\top\right]\mathbb{E}\left[v_{K_2,j}(\boldsymbol{X})v_{K_2,l}(\boldsymbol{X})\right](\lambda_l - \lambda_l^K) \quad \text{(by (13))} \\
&= \|\Lambda - \Lambda_{K_1\times K_2}\| \, \|\nabla G^*_{K_1\times K_2}(\Lambda_{K_1\times K_2})\| - \frac{a_3}{2\eta_2}\sum_{l=1}^{K_2}\sum_{j=1}^{K_2}(\lambda_j - \lambda_j^K)^\top \mathbb{E}\left[v_{K_2,j}(\boldsymbol{X})v_{K_2,l}(\boldsymbol{X})\right](\lambda_l - \lambda_l^K) \\
&= \|\Lambda - \Lambda_{K_1\times K_2}\| \, \|\nabla G^*_{K_1\times K_2}(\Lambda_{K_1\times K_2})\| - \frac{a_3}{2\eta_2}\sum_{j=1}^{K_2}(\lambda_j - \lambda_j^K)^\top(\lambda_j - \lambda_j^K) \\
&= \|\Lambda - \Lambda_{K_1\times K_2}\| \, \|\nabla G^*_{K_1\times K_2}(\Lambda_{K_1\times K_2})\| - \frac{a_3}{2\eta_2}\|\Lambda - \Lambda_{K_1\times K_2}\|^2 \\
&= \|\Lambda - \Lambda_{K_1\times K_2}\|\left(\|\nabla G^*_{K_1\times K_2}(\Lambda_{K_1\times K_2})\| - \frac{a_3}{2\eta_2}\|\Lambda - \Lambda_{K_1\times K_2}\|\right) \\
&\leq \|\Lambda - \Lambda_{K_1\times K_2}\|\left(\frac{aC}{\sqrt{\eta_1}}K^{-\alpha} - \frac{a_3}{2\eta_2}\cdot C_2K^{-\alpha}\right), \quad \text{(by (19))}
\end{aligned}$$

where $\bar\Lambda_{K_1\times K_2} = (\bar\lambda_1^K, ..., \bar\lambda_{K_2}^K)$ lies on the line joining $\Lambda = (\lambda_1, ..., \lambda_{K_2})$ and $\Lambda_{K_1\times K_2} =$

$(\lambda_1^K, ..., \lambda_{K_2}^K)$, which implies $u_{K_1}^\top(t)\bar{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x}) \in \Gamma_1$ by (16); $a_3 = \inf_{y\in\Gamma_1}\{\rho'(y)\} > 0$ is a finite positive constant; the fourth and fifth equalities follow from $\mathbb{E}\left[u_{K_1}(T)u_{K_1}(T)^\top\right] = I_{K_1\times K_1}$ and $\mathbb{E}\left[v_{K_2}(\boldsymbol{X})v_{K_2}(\boldsymbol{X})^\top\right] = I_{K_2\times K_2}$ respectively. Therefore, by choosing

$$C_2 > \frac{2\eta_2}{a_3}\cdot\frac{aC}{\sqrt{\eta_1}},$$

we can obtain the following conclusion:

$$G^*_{K_1\times K_2}(\Lambda_{K_1\times K_2}) > G^*_{K_1\times K_2}(\Lambda)\ , \quad \forall\Lambda\in\partial\Upsilon_{K_1\times K_2}\ . \tag{20}$$

Since $G^*_{K_1\times K_2}$ is continuous, (20) implies that there exists a local maximum of $G^*_{K_1\times K_2}$ in the interior of $\Upsilon_{K_1\times K_2}$. Note that $G^*_{K_1\times K_2}$ is strictly concave with a unique global maximum point $\Lambda^*_{K_1\times K_2}$, therefore we can claim that

$$\Lambda^*_{K_1\times K_2}\in\Upsilon^\circ_{K_1\times K_2},\ \text{i.e. } \|\Lambda^*_{K_1\times K_2}-\Lambda_{K_1\times K_2}\| = O(K^{-\alpha})\ . \tag{21}$$

By Mean Value Theorem, (16) and (21), we can deduce that

$$\begin{aligned}
&|\rho'\left(u_{K_1}(t)\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right) - \rho'\left(u_{K_1}(t)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)|\\
=&|\rho''(\xi^*(t,\boldsymbol{x}))|\left|u_{K_1}(t)\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{x}) - u_{K_1}(t)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right|\\
\le&-\rho''(\xi^*(t,\boldsymbol{x}))\times\|\Lambda_{K_1\times K_2}-\Lambda^*_{K_1\times K_2}\|\times\sup_{t\in\mathcal{T}}\|u_{K_1}(t)\|\times\sup_{\boldsymbol{x}\in\mathcal{X}}\|v_{K_2}(\boldsymbol{x})\|\\
\le&a_2C_2K^{-\alpha}\zeta_1(K_1)\zeta_2(K_2)\ ,
\end{aligned}$$

where $a_2 = \sup_{\gamma\in\Gamma_1}\{-\rho''(\gamma)\} < \infty$ is a finite positive constant, and $\xi^*(t,\boldsymbol{x})$ lies between the point $u_{K_1}(t)^\top\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})$ and $u_{K_1}(t)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{x})$ (note (16) implies $\xi^*(t,\boldsymbol{x})\in\Gamma_1$ for all $(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}$ and large enough K). Therefore, using the triangle inequality, and Assumption 1.4, we can have

$$\begin{aligned}
&\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}|\pi_0(t,\boldsymbol{x})-\pi^*_K(t,\boldsymbol{x})|\\
\le&\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}|\pi_0(t,\boldsymbol{x})-\rho'\left(u_{K_1}(t)\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)|\\
&+\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}\left|\rho'\left(u_{K_1}(t)\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)-\rho'\left(u_{K_1}(t)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)\right|\\
\le&\ aCK^{-\alpha}+a_2C_2K^{-\alpha}\zeta_1(K_1)\zeta_2(K_2) = O\left(K^{-\alpha}\zeta(K)\right),
\end{aligned}$$

where $\zeta(K) = \zeta_1(K_1)\zeta_2(K_2)$.

We next prove $\mathbb{E}\left[|\pi_0(T,\boldsymbol{X}) - \pi_K^*(T,\boldsymbol{X})|^2\right] = O\left(K^{-2\alpha}\right)$. By Assumption 1.4, we can deduce that

$$
\begin{aligned}
&\mathbb{E}\left[|\pi_0(T,\boldsymbol{X}) - \pi_K^*(T,\boldsymbol{X})|^2\right] \\
\leq& 2\cdot\mathbb{E}\left[|\pi_0(T,\boldsymbol{X}) - \rho'\left(u_{K_1}(T)\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)|^2\right] + 2\cdot\mathbb{E}\left[\left|\rho'\left(u_{K_1}(T)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right) - \rho'\left(u_{K_1}(T)\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)\right|^2\right] \\
\leq& 2\cdot\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}|\pi_0(t,\boldsymbol{x}) - \rho'\left(u_{K_1}(t)\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)|^2 + 2\sup_{\gamma\in\Gamma_1}|\rho''(\gamma)|^2\cdot\mathbb{E}\left[\left|u_{K_1}^\top(T)\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X})\right|^2\right] \\
\leq& O(K^{-2\alpha}) + O(1)\cdot\mathbb{E}\left[\left|u_{K_1}^\top(T)\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X})\right|^2\right].
\end{aligned}
$$

We next compute the order of $\mathbb{E}\left[\left|u_{K_1}^\top(T)\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X})\right|^2\right]$. Note that $\mathbb{E}[u_{K_1}(T)u_{K_1}(T)^\top] = I_{K_1\times K_1}$, $\mathbb{E}[v_{K_2}(\boldsymbol{X})v_{K_2}(\boldsymbol{X})^\top] = I_{K_2\times K_2}$, (13), (21) and Assumption 1.2, we can deduce that

$$
\begin{aligned}
&\mathbb{E}\left[\left|u_{K_1}^\top(T)\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X})\right|^2\right] \\
=&\mathbb{E}\left[u_{K_1}^\top(T)\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X})v_{K_2}(\boldsymbol{X})^\top\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}^\top u_{K_1}(T)\right] \\
=&\mathbb{E}\left[\frac{1}{\pi_0(T,\boldsymbol{X})}\pi_0(T,\boldsymbol{X})u_{K_1}^\top(T)\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X})v_{K_2}(\boldsymbol{X})^\top\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}^\top u_{K_1}(T)\right] \\
\leq&\frac{1}{\eta_1}\cdot\mathbb{E}\left[\pi_0(T,\boldsymbol{X})u_{K_1}^\top(T)\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X})v_{K_2}(\boldsymbol{X})^\top\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}^\top u_{K_1}(T)\right] \\
=&\frac{1}{\eta_1}\cdot\int_{\mathcal{T}}u_{K_1}^\top(t)\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}\mathbb{E}\left[v_{K_2}(\boldsymbol{X})v_{K_2}(\boldsymbol{X})^\top\right]\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}^\top u_{K_1}(t)dF_T(t) \quad \text{(by (13))} \\
=&\frac{1}{\eta_1}\cdot\int_{\mathcal{T}}u_{K_1}^\top(t)\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}\cdot\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}^\top u_{K_1}(t)dF_T(t) \\
=&\frac{1}{\eta_1}\cdot\int_{\mathcal{T}}\operatorname{tr}\Bigg(\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}\cdot\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}^\top u_{K_1}(t)u_{K_1}^\top(t)\Bigg)dF_T(t) \\
=&\frac{1}{\eta_1}\cdot\operatorname{tr}\Bigg(\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}\cdot\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}^\top\Bigg) \\
\leq&\frac{1}{\eta_1}\cdot\|\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\|^2 = O(K^{-2\alpha}). \quad \text{(by (21))}
\end{aligned} \tag{22}
$$

Therefore, we can obtain

$$
\mathbb{E}\left[|\pi_0(T,\boldsymbol{X}) - \pi_K^*(T,\boldsymbol{X})|^2\right] = O\left(K^{-2\alpha}\right).
$$

We finally prove $N^{-1}\sum_{i=1}^N|\pi_0(T_i,\boldsymbol{X}_i) - \pi_K^*(T_i,\boldsymbol{X}_i)|^2 = O_p\left(K^{-2\alpha}\right)$. Note that by (22), we can have

$$
\mathbb{E}\left[\left\{\frac{1}{N}\sum_{i=1}^N\left|u_{K_1}^\top(T_i)\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X}_i)\right|^2 - \mathbb{E}\left[\left|u_{K_1}^\top(T)\left\{\Lambda^*_{K_1\times K_2} - \Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X})\right|^2\right]\right\}^2\right]
$$

$$
\begin{aligned}
&\le \frac{1}{N}\cdot\mathbb{E}\left[\left|u_{K_1}^\top(T)\left\{\Lambda^*_{K_1\times K_2}-\Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X})\right|^4\right]\\
&\le \frac{1}{N}\cdot\mathbb{E}\left[\left|u_{K_1}^\top(T)\left\{\Lambda^*_{K_1\times K_2}-\Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X})\right|^2\right]\cdot\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}\left|u_{K_1}^\top(t)\left\{\Lambda^*_{K_1\times K_2}-\Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{x})\right|^2\\
&\le \frac{1}{N}\cdot O(K^{-2\alpha})\cdot\zeta_1(K_1)^2\zeta_2(K_2)^2\cdot\left\|\Lambda^*_{K_1\times K_2}-\Lambda_{K_1\times K_2}\right\|^2\le\frac{1}{N}\cdot\zeta(K)^2\cdot O(K^{-4\alpha}),
\end{aligned}
$$

then in light of Chebyshev's inequality and Assumption 1.4, we have

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N\left|u_{K_1}^\top(T_i)\left\{\Lambda^*_{K_1\times K_2}-\Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X}_i)\right|^2-\mathbb{E}\left[\left|u_{K_1}^\top(T)\left\{\Lambda^*_{K_1\times K_2}-\Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X})\right|^2\right]\\
=&O_p\left(\frac{\zeta(K)}{\sqrt{N}}K^{-2\alpha}\right)=o_p\left(K^{-2\alpha}\right). \qquad (23)
\end{aligned}
$$

With (21), (22), (23), and Assumption 1.2, we can deduce that

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N\left|\pi_0(T_i,\boldsymbol{X}_i)-\pi_K^*(T_i,\boldsymbol{X}_i)\right|^2\\
\le&\frac{2}{N}\sum_{i=1}^N\left|\pi_0(T_i,\boldsymbol{X}_i)-\rho'\left(u_{K_1}(T_i)\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)\right)\right|^2\\
&+\frac{2}{N}\sum_{i=1}^N\left|\rho'\left(u_{K_1}(T_i)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)\right)-\rho'\left(u_{K_1}(T_i)\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)\right)\right|^2\\
\le&2\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}\left|\pi_0(t,\boldsymbol{x})-\rho'\left(u_{K_1}(t)\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)\right|^2\\
&+\sup_{\gamma\in\Gamma_1}|\rho''(\gamma)|^2\cdot\frac{2}{N}\sum_{i=1}^N\left|u_{K_1}^\top(T_i)\left\{\Lambda^*_{K_1\times K_2}-\Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X}_i)\right|^2\\
\le&2\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}\left|\pi_0(t,\boldsymbol{x})-\rho'\left(u_{K_1}(t)\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)\right|^2\\
&+2\cdot\sup_{\gamma\in\Gamma_1}|\rho''(\gamma)|^2\cdot\mathbb{E}\left[\left|u_{K_1}^\top(T)\left\{\Lambda^*_{K_1\times K_2}-\Lambda_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X})\right|^2\right]+o_p\left(K^{-2\alpha}\right)\\
=&O(K^{-2\alpha})+O(K^{-2\alpha})+o_p\left(K^{-2\alpha}\right)=O_p\left(K^{-2\alpha}\right). \quad \text{(by (21))}
\end{aligned}
$$

■

3.2 Lemma 3.2

Lemma 3.2 *Under Assumption 1.2-1.4, we have*

$$
\left\|\hat{\Lambda}_{K_1\times K_2}-\Lambda^*_{K_1\times K_2}\right\|=O_p\left(\sqrt{\frac{K}{N}}\right).
$$

Proof. Define

$$\hat{S}_N := \frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{K_2}\sum_{j=1}^{K_2}(\lambda_j-\lambda_j^*)^\top \pi_0(T_i,\boldsymbol{X}_i)u_{K_1}(T_i)u_{K_1}(T_i)^\top(\lambda_l-\lambda_l^*)v_{K_2,j}(\boldsymbol{X}_i)v_{K_2,l}(\boldsymbol{X}_i),$$

where λ_j and λ_j^* are the j-th column of Λ and $\Lambda^*_{K_1\times K_2}$ respectively. Since $\hat{S}_N$ is symmetric, using (13) and the facts that $\mathbb{E}\left[u_{K_1}(T)u_{K_1}(T)^\top\right]=I_{K_1\times K_1}$ and $\mathbb{E}\left[v_{K_2}(\boldsymbol{X})v_{K_2}(\boldsymbol{X})^\top\right]=I_{K_2\times K_2}$, we can have

$$\begin{aligned}
\mathbb{E}\left[\hat{S}_N\right] &= \sum_{l=1}^{K_2}\sum_{j=1}^{K_2}(\lambda_j-\lambda_j^*)^\top \mathbb{E}\left[\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^\top v_{K_2,j}(\boldsymbol{X})v_{K_2,l}(\boldsymbol{X})\right](\lambda_l-\lambda_l^*)\\
&= \sum_{l=1}^{K_2}\sum_{j=1}^{K_2}(\lambda_j-\lambda_j^*)^\top \mathbb{E}\left[u_{K_1}(T)u_{K_1}(T)^\top\right]\mathbb{E}[v_{K_2,j}(\boldsymbol{X})v_{K_2,l}(\boldsymbol{X})](\lambda_l-\lambda_l^*)\\
&= \sum_{j=1}^{K_2}(\lambda_j-\lambda_j^*)^\top(\lambda_j-\lambda_j^*) = \left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|.
\end{aligned}$$

Then we can further deduce that

$$\begin{aligned}
&\mathbb{E}\left[\left|\hat{S}_N-\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|\right|^2\right]\\
=&\mathbb{E}[\hat{S}_N^2]-2\mathbb{E}[\hat{S}_N]\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|+\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|^2\\
=&\frac{N}{N^2}\cdot\mathbb{E}\left[\left(\sum_{l=1}^{K_2}\sum_{j=1}^{K_2}(\lambda_j-\lambda_j^*)^\top\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^\top(\lambda_l-\lambda_l^*)v_{K_2,j}(\boldsymbol{X})v_{K_2,l}(\boldsymbol{X})\right)^2\right]\\
&+2\cdot\frac{1}{N^2}\cdot\binom{N}{2}\cdot\mathbb{E}\left[\sum_{l=1}^{K_2}\sum_{j=1}^{K_2}(\lambda_j-\lambda_j^*)^\top\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^\top(\lambda_l-\lambda_l^*)v_{K_2,j}(\boldsymbol{X})v_{K_2,l}(\boldsymbol{X})\right]^2\\
&-\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|^2\\
=&\frac{1}{N}\mathbb{E}\left[\left(\sum_{l=1}^{K_2}\sum_{j=1}^{K_2}(\lambda_j-\lambda_j^*)^\top\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^\top(\lambda_l-\lambda_l^*)v_{K_2,j}(\boldsymbol{X})v_{K_2,l}(\boldsymbol{X})\right)^2\right]\\
&+\frac{N(N-1)}{N^2}\cdot\mathbb{E}\left[\hat{S}_N\right]^2-\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|^2\\
=&\frac{1}{N}\mathbb{E}\left[\left(\sum_{l=1}^{K_2}\sum_{j=1}^{K_2}(\lambda_j-\lambda_j^*)^\top\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^\top(\lambda_l-\lambda_l^*)v_{K_2,j}(\boldsymbol{X})v_{K_2,l}(\boldsymbol{X})\right)^2\right]\\
&-\frac{1}{N}\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|^2
\end{aligned}$$

$$<\frac{1}{N}\mathbb{E}\left[\left(\sum_{l=1}^{K_2}\sum_{j=1}^{K_2}(\lambda_j-\lambda_j^*)^\top\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^\top(\lambda_l-\lambda_l^*)v_{K_2,j}(\boldsymbol{X})v_{K_2,l}(\boldsymbol{X})\right)^2\right].$$

In light of the fact that

$$0\le y^\top\left\{\pi_0(t,\boldsymbol{x})u_{K_1}(t)u_{K_1}(t)^\top\right\}y\le\eta_2\zeta_1(K_1)^2y^\top y\ ,\ \forall y\in\mathbb{R}^{K_1},\ \forall(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}\ ,$$

we can deduce that

$$\begin{aligned}&\sum_{l=1}^{K_2}\sum_{j=1}^{K_2}(\lambda_j-\lambda_j^*)^\top\left\{\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^\top\right\}(\lambda_l-\lambda_l^*)v_{K_2,j}(\boldsymbol{X})v_{K_2,l}(\boldsymbol{X})\\
=&\left[\sum_{j=1}^{K_2}v_{K_2,j}(\boldsymbol{X})(\lambda_j-\lambda_j^*)^\top\right]\cdot\left\{\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^\top\right\}\cdot\left[\sum_{l=1}^{K_2}(\lambda_l-\lambda_l^*)v_{K_2,l}(\boldsymbol{X})\right]\\
\le&\eta_2\cdot\|u_{K_1}(T)\|^2\cdot\left\|\sum_{i=1}^{K_2}(\lambda_i-\lambda_i^*)^\top v_{K_2,i}(\boldsymbol{X})\right\|^2\\
\le&\eta_2\cdot\|u_{K_1}(T)\|^2\cdot\left(\sum_{i=1}^{K_2}\|\lambda_i-\lambda_i^*\|^2\right)\left(\sum_{i=1}^{K_2}v_{K_2,i}(\boldsymbol{X})^2\right)\\
=&\eta_2\cdot\|u_{K_1}(T)\|^2\cdot\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|^2\|v_{K_2}(\boldsymbol{X})\|^2.\end{aligned}$$

Therefore, we can obtain that

$$\begin{aligned}&\mathbb{E}\left[\left|\hat{S}_N-\|\Lambda-\Lambda^*_{K_1\times K_2}\|\right|^2\right]\\
\le&\frac{1}{N}\eta_2^2\cdot\mathbb{E}\left[\|u_{K_1}(T)\|^4\cdot\|v_{K_2}(\boldsymbol{X})\|^4\right]\cdot\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|^4\\
\le&\frac{1}{N}\eta_2^2\cdot\zeta_1(K_1)^2\cdot\zeta_2(K_2)^2\cdot\mathbb{E}\left[\|u_{K_1}(T)\|^2\cdot\|v_{K_2}(\boldsymbol{X})\|^2\right]\cdot\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|^4\\
=&\frac{1}{N}\eta_2^2\cdot\zeta_1(K_1)^2\cdot\zeta_2(K_2)^2\cdot\mathbb{E}\left[\frac{1}{\pi_0(T,\boldsymbol{X})}\cdot\pi_0(T,\boldsymbol{X})\|u_{K_1}(T)\|^2\cdot\|v_{K_2}(\boldsymbol{X})\|^2\right]\cdot\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|^4\\
\le&\frac{1}{N}\frac{\eta_2^2}{\eta_1}\cdot\zeta_1(K_1)^2\cdot\zeta_2(K_2)^2\cdot\mathbb{E}\left[\pi_0(T,\boldsymbol{X})\|u_{K_1}(T)\|^2\cdot\|v_{K_2}(\boldsymbol{X})\|^2\right]\cdot\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|^4\quad\text{(by Assumption 1.2)}\\
=&\frac{1}{N}\frac{\eta_2^2}{\eta_1}\cdot\zeta_1(K_1)^2\cdot\zeta_2(K_2)^2\cdot\mathbb{E}\left[\|u_{K_1}(T)\|^2\right]\cdot\mathbb{E}\left[\|v_{K_2}(\boldsymbol{X})\|^2\right]\cdot\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|^4\quad\text{(by (13))}\\
=&\frac{1}{N}\frac{\eta_2^2}{\eta_1}\cdot\zeta_1(K_1)^2\cdot\zeta_2(K_2)^2\cdot K_1\cdot K_2\cdot\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|^4\text{(since }\mathbb{E}[\|u_{K_1}(T)\|^2]=K_1\text{ and }\mathbb{E}[\|v_{K_2}(\boldsymbol{X})\|^2]=K_2)\\
=&\frac{1}{N}\frac{\eta_2^2}{\eta_1}\cdot\zeta(K)^2\cdot K\cdot\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|^4\ .\quad\text{(since }\zeta(K)=\zeta_1(K_1)\zeta_2(K_2)\text{ and }K=K_1\cdot K_2)\end{aligned}\tag{24}$$

Considering the event set

$$E_N := \left\{ \hat{S}_N > \frac{1}{2} \left\| \Lambda - \Lambda^*_{K_1\times K_2} \right\|^2 ,\ \Lambda \neq \Lambda^*_{K_1\times K_2} \right\} \ ,$$

by Chebyshev's inequality, (24), and Assumption 1.4 we can get

$$\begin{aligned}
&\mathbb{P}\left(\left| \hat{S}_N - \left\| \Lambda - \Lambda^*_{K_1\times K_2} \right\|^2 \right| \geq \frac{1}{2} \left\| \Lambda - \Lambda^*_{K_1\times K_2} \right\|^2 ,\ \Lambda \neq \Lambda^*_{K_1\times K_2} \right) \\
&\leq \frac{4\mathbb{E}\left[\left| \hat{S}_N - \left\| \Lambda - \Lambda^*_{K_1\times K_2} \right\| \right|^2 \right]}{\left\| \Lambda - \Lambda^*_{K_1\times K_2} \right\|^4} \\
&\leq \frac{4}{N} \frac{\eta_2^2}{\eta_1} \cdot \zeta(K)^2 \cdot K \leq O\left(\frac{\zeta(K)^2 K}{N} \right) = o(1).
\end{aligned} \tag{25}$$

Note that

$$\begin{aligned}
&\nabla \hat{G}_{K_1\times K_2}(\Lambda) = \frac{1}{N}\sum_{i=1}^{N} \left\{ \rho'\left(u_{K_1}^\top(T_i)\Lambda v_{K_2}(\boldsymbol{X}_i) \right) u_{K_1}(T_i) v_{K_2}^\top(\boldsymbol{X}_i) - u_{K_1}(T_i) \cdot \mathbb{E}[v_{K_2}^\top(\boldsymbol{X})] \right\} \\
&- \mathbb{E}[u_{K_1}(T)] \cdot \left\{ \frac{1}{N}\sum_{l=1}^{N} v_{K_2}^\top(\boldsymbol{X}_l) - \mathbb{E}[v_{K_2}^\top(\boldsymbol{X})] \right\} \\
&- \left\{ \frac{1}{N}\sum_{i=1}^{N} u_{K_1}(T_i) - \mathbb{E}[u_{K_1}(T)] \right\} \left\{ \frac{1}{N}\sum_{l=1}^{N} v_{K_2}^\top(\boldsymbol{X}_l) - \mathbb{E}[v_{K_2}^\top(\boldsymbol{X})] \right\} \\
=&\nabla \widehat{H}_{K_1\times K_2}(\Lambda) - \left\{ \frac{1}{N}\sum_{i=1}^{N} u_{K_1}(T_i) - \mathbb{E}[u_{K_1}(T)] \right\} \left\{ \frac{1}{N}\sum_{l=1}^{N} v_{K_2}^\top(\boldsymbol{X}_l) - \mathbb{E}[v_{K_2}^\top(\boldsymbol{X})] \right\},
\end{aligned} \tag{26}$$

where

$$\begin{aligned}
\widehat{H}_{K_1\times K_2}(\Lambda) := &\frac{1}{N}\sum_{i=1}^{N} \left\{ \rho\left(u_{K_1}^\top(T_i)\Lambda v_{K_2}(\boldsymbol{X}_i) \right) - u_{K_1}(T_i)^\top \Lambda \mathbb{E}[v_{K_2}(\boldsymbol{X})] \right\} \\
&- \mathbb{E}[u_{K_1}^\top(T)]\Lambda \left\{ \frac{1}{N}\sum_{l=1}^{N} v_{K_2}(\boldsymbol{X}_l) - \mathbb{E}[v_{K_2}(\boldsymbol{X})] \right\}.
\end{aligned}$$

Since $\Lambda^*_{K_1\times K_2}$ is a unique maximizer of $G^*_{K_1\times K_2}(\cdot)$, then for each $j \in \{1, \ldots, K_2\}$,

$$\begin{aligned}
&\frac{\partial}{\partial \lambda_j} G^*_{K_1\times K_2}(\lambda_1^*, \ldots, \lambda_{K_2}^*) \\
=&\mathbb{E}\left[\rho'\left(u_{K_1}^\top(T)\Lambda^*_{K_1\times K_2} v_{K_2}(\boldsymbol{X}) \right) u_{K_1}(X) v_{K_2,j}(Y) \right] - \mathbb{E}[u_{K_1}(T)]\mathbb{E}[v_{K_2,j}(\boldsymbol{X})] = 0 \ .
\end{aligned}$$

Therefore, for large enough K, we can deduce that

$$
\begin{aligned}
&\mathbb{E}\left[\|\nabla \hat{H}_{K_1\times K_2}(\Lambda^*_{K_1\times K_2})\|^2\right] = \sum_{j=1}^{K_2}\mathbb{E}\left[\left\|\frac{\partial}{\partial\lambda_j}\hat{H}_{K_1\times K_2}(\lambda_1^*,\ldots,\lambda_{K_2}^*)\right\|^2\right] \qquad (27)\\
\leq & 2\sum_{j=1}^{K_2}\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\left\{\rho'\left(u_{K_1}^\top(T_i)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)\right)u_{K_1}(T_i)v_{K_2,j}(\boldsymbol{X}_i)-\mathbb{E}[v_{K_2,j}(\boldsymbol{X})]u_{K_1}(T_i)\right\}\right\|^2\right]\\
&+2\sum_{j=1}^{K_2}\mathbb{E}\left[\left\|\mathbb{E}\left[u_{K_1}(T)\right]\cdot\left\{\frac{1}{N}\sum_{l=1}^{N}v_{K_2,j}(\boldsymbol{X}_l)-\mathbb{E}[v_{K_2,j}(\boldsymbol{X})]\right\}\right\|^2\right]\\
\leq & \frac{4}{N}\sum_{j=1}^{K_2}\left\{\mathbb{E}\left[\left\|\rho'\left(u_{K_1}^\top(T)\Lambda^*_{K_1\times K_2}v_{K_2}(Y)\right)u_{K_1}(T)v_{K_2,j}(\boldsymbol{X})\right\|^2\right]+\mathbb{E}[v_{K_2,j}(\boldsymbol{X})^2]\mathbb{E}\left[\|u_{K_1}(T)\|^2\right]\right\}\\
&+\frac{2}{N}\sum_{j=1}^{K_2}\mathbb{E}\left[v_{K_2,j}(\boldsymbol{X})^2\right]\cdot\mathbb{E}\left[\|u_{K_1}(T)\|^2\right]\\
= & \frac{4}{N}\sum_{j=1}^{K_2}\left\{\mathbb{E}\left[\frac{\left|\rho'\left(u_{K_1}^\top(T)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{X})\right)\right|^2}{\pi_0(T,\boldsymbol{X})}\cdot\pi_0(T,\boldsymbol{X})\cdot\|u_{K_1}(T)v_{K_2,j}(\boldsymbol{X})\|^2\right]+\mathbb{E}[v_{K_2,j}(\boldsymbol{X})^2]\mathbb{E}\left[\|u_{K_1}(T)\|^2\right]\right\}\\
&+\frac{2}{N}\sum_{j=1}^{K_2}\mathbb{E}\left[v_{K_2,j}(\boldsymbol{X})^2\right]\cdot\mathbb{E}\left[\|u_{K_1}(T)\|^2\right]\\
\leq & \frac{4}{N}\sum_{j=1}^{K_2}\left\{\frac{\left(\sup_{\gamma\in\Gamma_1}\rho'(\gamma)\right)^2}{\eta_1}\cdot\mathbb{E}\left[\pi_0(T,\boldsymbol{X})\cdot\|u_{K_1}(T)v_{K_2,j}(\boldsymbol{X})\|^2\right]+\mathbb{E}[v_{K_2,j}(\boldsymbol{X})^2]\mathbb{E}\left[\|u_{K_1}(T)\|^2\right]\right\}\\
&+\frac{2}{N}\sum_{j=1}^{K_2}\mathbb{E}\left[v_{K_2,j}(\boldsymbol{X})^2\right]\cdot\mathbb{E}\left[\|u_{K_1}(T)\|^2\right]\\
= & \frac{4}{N}\sum_{j=1}^{K_2}\left\{\frac{\left(\sup_{\gamma\in\Gamma_1}\rho'(\gamma)\right)^2}{\eta_1}\cdot\mathbb{E}\left[v_{K_2,j}(\boldsymbol{X})^2\right]\mathbb{E}\left[\|u_{K_1}(T)\|^2\right]+\mathbb{E}[v_{K_2,j}(\boldsymbol{X})^2]\mathbb{E}\left[\|u_{K_1}(T)\|^2\right]\right\}\\
&+\frac{2}{N}\sum_{j=1}^{K_2}\mathbb{E}\left[v_{K_2,j}(\boldsymbol{X})^2\right]\cdot\mathbb{E}\left[\|u_{K_1}(T)\|^2\right]\\
\leq & \frac{1}{N}\left\{\frac{4}{\eta_1}\left(\sup_{\gamma\in\Gamma_1}\rho'(\gamma)\right)^2+4+2\right\}\cdot\mathbb{E}\left[\|u_{K_1}(T)\|^2\right]\sum_{j=1}^{K_2}\mathbb{E}\left[v_{K_2,j}(\boldsymbol{X})^2\right]\\
= & \frac{1}{N}\left\{\frac{4}{\eta_1}\left(\sup_{\gamma\in\Gamma_1}\rho'(\gamma)\right)^2+6\right\}K_1K_2\leq C_4^2\frac{K}{N}\ ,
\end{aligned}
$$

where the last inequality follows by Assumption 1.8 and C_4 is a finite universal constant.

Let $\epsilon > 0$, fix $C_5(\epsilon) > 0$ (to be chosen later) and define

$$\hat{\Upsilon}_{K_1\times K_2}(\epsilon) := \left\{ \Lambda \in \mathbb{R}^{K_1\times K_2} : \|\Lambda - \Lambda^*_{K_1\times K_2}\| \le C_5(\epsilon) C_4 \sqrt{\frac{K}{N}} \right\}.$$

For $\forall \Lambda \in \hat{\Upsilon}_{K_1\times K_2}(\epsilon), \forall (t, \boldsymbol{x}) \in \mathcal{T} \times \mathcal{X}$, we can have

$$\begin{aligned}
&\left| u_{K_1}(t)^\top \Lambda v_{K_2}(\boldsymbol{x}) - u_{K_1}(t)^\top \Lambda^*_{K_1\times K_2} v_{K_2}(\boldsymbol{x}) \right| \\
\le& \|\Lambda - \Lambda^*_{K_1\times K_2}\| \sup_{t\in\mathcal{T}} \|u_{K_1}(t)\| \sup_{\boldsymbol{x}\in\mathcal{X}} \|v_{K_2}(\boldsymbol{x})\| \le C_5(\epsilon) C_4 \sqrt{\frac{K}{N}} \zeta_1(K_1)\zeta_2(K_2),
\end{aligned}$$

thus for large enough N, in accordance with Assumption 1.4 and (14), we have

$$\begin{aligned}
u_{K_1}(t)^\top \Lambda v_{K_2}(\boldsymbol{x}) \in & \left[u_{K_1}(t)^\top \Lambda^*_{K_1\times K_2} v_{K_2}(\boldsymbol{x}) - C_5(\epsilon) C_4 \zeta_1(K_1)\zeta_2(K_2) \sqrt{\frac{K}{N}}, \right. \\
& \qquad \left. u_{K_1}(t)^\top \Lambda^*_{K_1\times K_2} v_{K_2}(\boldsymbol{x}) + C_5(\epsilon) C_4 \zeta_1(K_1)\zeta_2(K_2) \sqrt{\frac{K}{N}} \right] \\
& \subset \left[\underline{\gamma} - C K^{-\alpha} - C_5(\epsilon) C_4 \zeta_1(K_1)\zeta_2(K_2) \sqrt{\frac{K}{N}}, \right. \\
& \qquad \left. \overline{\gamma} + C K^{-\alpha} + C_5(\epsilon) C_4 \zeta_1(K_1)\zeta_2(K_2) \sqrt{\frac{K}{N}} \right] \subset \Gamma_2(\epsilon) \ ,
\end{aligned} \tag{28}$$

where $\Gamma_2(\epsilon) := \left[\underline{\gamma} - 1 - C_5(\epsilon), \overline{\gamma} + 1 + C_5(\epsilon) \right]$ is a compact set and independent of $(t, \boldsymbol{x})$.

For any $\Lambda \in \partial \hat{\Upsilon}_{K_1\times K_2}(\epsilon)$, there exists $\bar{\Lambda}$ on the line joining Λ and $\Lambda^*_{K_1\times K_2}$ such that

$$\begin{aligned}
\hat{G}_{K_1\times K_2}(\Lambda) =& \hat{G}_{K_1\times K_2}(\Lambda^*_{K_1\times K_2}) + \sum_{j=1}^{K_2} (\lambda_j - \lambda_j^*)^\top \frac{\partial}{\partial \lambda_i} \hat{G}_{K_1\times K_2}(\lambda_1^*, \ldots, \lambda_{K_2}^*) \\
&+ \frac{1}{2} \sum_{l=1}^{K_2} \sum_{j=1}^{K_2} (\lambda_j - \lambda_j^*)^\top \frac{\partial^2}{\partial \lambda_i \partial \lambda_l} \hat{G}_{K_1\times K_2}(\bar{\lambda}_1, \ldots, \bar{\lambda}_{K_2})(\lambda_l - \lambda_l^*) \ ,
\end{aligned}$$

where $\bar{\lambda}_j$ denotes the j-th column of $\bar{\Lambda}$. For the second order term in above equality, note that $u_{K_1}^\top(t)\bar{\Lambda} v_{K_2}(\boldsymbol{x}) \in \Gamma_2(\epsilon)$ for all $(t, \boldsymbol{x}) \in \mathcal{T} \times \mathcal{X}$, we can further deduce that

$$\sum_{l=1}^{K_2} \sum_{j=1}^{K_2} (\lambda_j - \lambda_j^*)^\top \frac{\partial^2}{\partial \lambda_i \partial \lambda_l} \hat{G}_{K_1\times K_2}(\bar{\lambda}_1, \ldots, \bar{\lambda}_{K_2})(\lambda_l - \lambda_l^*) \tag{29}$$

$$
\begin{aligned}
=&\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K_2}\sum_{l=1}^{K_2}(\lambda_j-\lambda_j^*)^\top u_{K_1}(T_i)\rho''\left(u_{K_1}^\top(T_i)\bar{\Lambda}v_{K_2}(\boldsymbol{X}_i)\right)(\lambda_l-\lambda_l^*)^\top u_{K_1}(T_i)v_{K_2,j}(\boldsymbol{X}_i)v_{K_2,l}(\boldsymbol{X}_i)\\
\leq&-\frac{\bar{b}(\epsilon)}{N}\sum_{i=1}^{N}\sum_{j=1}^{K_2}\sum_{l=1}^{K_2}(\lambda_j-\lambda_j^*)^\top u_{K_1}(T_i)u_{K_1}(T_i)^\top(\lambda_l-\lambda_l^*)v_{K_2,j}(\boldsymbol{X}_i)v_{K_2,l}(\boldsymbol{X}_i)\\
=&-\frac{\bar{b}(\epsilon)}{N}\sum_{i=1}^{N}\sum_{j=1}^{K_2}\sum_{l=1}^{K_2}\frac{1}{\pi_0(T_i,\boldsymbol{X}_i)}(\lambda_j-\lambda_j^*)^\top\pi_0(T_i,\boldsymbol{X}_i)u_{K_1}(T_i)u_{K_1}(T_i)^\top(\lambda_l-\lambda_l^*)v_{K_2,j}(\boldsymbol{X}_i)v_{K_2,l}(\boldsymbol{X}_i)\\
\leq&-\frac{\bar{b}(\epsilon)}{N\eta_2}\sum_{i=1}^{N}\sum_{j=1}^{K_2}\sum_{l=1}^{K_2}(\lambda_j-\lambda_j^*)^\top\pi_0(T_i,\boldsymbol{X}_i)u_{K_1}(T_i)u_{K_1}(T_i)^\top(\lambda_l-\lambda_l^*)v_{K_2,j}(\boldsymbol{X}_i)v_{K_2,l}(\boldsymbol{X}_i)\\
=&-\frac{\bar{b}(\epsilon)}{\eta_2}\hat{S}_N\ ,
\end{aligned}
$$

where $-\bar{b}(\epsilon):=\sup_{\gamma\in\Gamma_2(\epsilon)}\rho''(\gamma)<\infty$.

Define the event set

$$
\begin{aligned}
E_N:=\Bigg\{&\hat{S}_N>\frac{1}{2}\left\|\Lambda-\Lambda_{K_1\times K_2}^*\right\|^2,\ \Lambda\neq\Lambda_{K_1\times K_2}^*;\\
&\text{and}\ \left\|\left\{\frac{1}{N}\sum_{i=1}^{N}u_{K_1}(T_i)-\mathbb{E}[u_{K_1}(T)]\right\}\left\{\frac{1}{N}\sum_{l=1}^{N}v_{K_2}^\top(\boldsymbol{X}_l)-\mathbb{E}[v_{K_2}^\top(\boldsymbol{X})]\right\}\right\|\leq\frac{1}{N^{1/4}}\cdot\sqrt{\frac{K}{N}}\Bigg\}.
\end{aligned}
$$

Note that

$$
\left\|\left\{\frac{1}{N}\sum_{i=1}^{N}u_{K_1}(T_i)-\mathbb{E}[u_{K_1}(T)]\right\}\left\{\frac{1}{N}\sum_{l=1}^{N}v_{K_2}^\top(\boldsymbol{X}_l)-\mathbb{E}[v_{K_2}^\top(\boldsymbol{X})]\right\}\right\|=O_P\left(\frac{\sqrt{K}}{N}\right).
$$

By (25), we can deduece that for any $\epsilon>0$, there exists $N_0(\epsilon)\in\mathbb{N}$ such that $N>N_0(\epsilon)$ large enough

$$
\begin{aligned}
&\mathbb{P}\left((E_N)^c\right)<\mathbb{P}\left(\left|\hat{S}_N-\left\|\Lambda-\Lambda_{K_1\times K_2}^*\right\|^2\right|\geq\frac{1}{2}\left\|\Lambda-\Lambda_{K_1\times K_2}^*\right\|^2,\ \Lambda\neq\Lambda_{K_1\times K_2}^*\right)\\
&\quad+\mathbb{P}\left(\left\|\left\{\frac{1}{N}\sum_{i=1}^{N}u_{K_1}(T_i)-\mathbb{E}[u_{K_1}(T)]\right\}\left\{\frac{1}{N}\sum_{l=1}^{N}v_{K_2}^\top(\boldsymbol{X}_l)-\mathbb{E}[v_{K_2}^\top(\boldsymbol{X})]\right\}\right\|>\frac{1}{N^{1/4}}\cdot\sqrt{\frac{K}{N}}\right)\\
<&\frac{\epsilon}{4}+\frac{\epsilon}{4}=\frac{\epsilon}{2}\ . \qquad (30)
\end{aligned}
$$

Therefore, **on the event** E_N, for large enough N, we can deduce that for any $\Lambda\in$

$\partial\hat{\Upsilon}_{K_1\times K_2}(\epsilon)$,

$$\begin{aligned}
&\hat{G}_{K_1\times K_2}(\Lambda)-\hat{G}_{K_1\times K_2}(\Lambda^*_{K_1\times K_2}) \qquad (31)\\
=&\sum_{j=1}^{K_2}(\lambda_j-\lambda_j^*)^\top\frac{\partial}{\partial\lambda_j}\hat{G}_{K_1\times K_2}(\lambda_1^*,\ldots,\lambda_{K_2}^*)\\
&+\sum_{l=1}^{K_2}\sum_{j=1}^{K_2}\frac{1}{2}(\lambda_j-\lambda_j^*)^\top\frac{\partial^2}{\partial\lambda_j\partial\lambda_l}\hat{G}_{K_1\times K_2}(\bar{\lambda}_1,\ldots,\bar{\lambda}_{K_2})(\lambda_l-\lambda_l^*)\\
\leq&\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|\|\nabla\hat{G}_{K_1\times K_2}(\Lambda^*_{K_1\times K_2})\|-\frac{\bar{b}(\epsilon)}{2\eta_2}\hat{S}_N \text{ (by (29))}\\
\leq&\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|\|\nabla\hat{G}_{K_1\times K_2}(\Lambda^*_{K_1\times K_2})\|-\frac{\bar{b}(\epsilon)}{4\eta_2}\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|^2\\
\leq&\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|\left(\|\nabla\hat{H}_{K_1\times K_2}(\Lambda^*_{K_1\times K_2})\|+\frac{1}{N^{1/4}}\cdot\sqrt{\frac{K}{N}}-\frac{\bar{b}(\epsilon)}{4\eta_2}\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|\right)\quad\text{(by (26))}\\
\leq&\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|\left(\|\nabla\hat{H}_{K_1\times K_2}(\Lambda^*_{K_1\times K_2})\|-\frac{1}{2}\cdot\frac{\bar{b}(\epsilon)}{4\eta_2}\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|\right)
\end{aligned}$$

where the second and the last inequality follow from definition of the event E_N.

Note that for sufficiently large N, by Chebyshev's inequality and (27) we have

$$\begin{aligned}
&\mathbb{P}\left\{\|\nabla\hat{H}_{K_1\times K_2}(\Lambda^*_{K_1\times K_2})\|\geq\frac{\bar{b}(\epsilon)}{8\eta_2}\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|\right\} \qquad (32)\\
\leq&\frac{64\cdot\eta_2^2}{\bar{b}(\epsilon)^2}\cdot\frac{\mathbb{E}\left[\left\|\nabla\hat{H}_{K_1\times K_2}(\Lambda^*_{K_1\times K_2})\right\|^2\right]}{\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|^2}\leq\frac{64\eta_2^2}{\bar{b}(\epsilon)^2C_5^2(\epsilon)}\leq\frac{\epsilon}{2}
\end{aligned}$$

where the last inequality holds by choosing

$$C_5(\epsilon)\geq\sqrt{\frac{128\cdot\eta_2^2}{\bar{b}(\epsilon)^2\epsilon}}\ .$$

Therefore, for sufficiently large N, by (30) and (32) we can derive

$$\begin{aligned}
&\mathbb{P}\left((E_N)^c\text{ or }\|\nabla\hat{H}_{K_1\times K_2}(\Lambda^*_{K_1\times K_2})\|\geq\frac{\bar{b}(\epsilon)}{8\eta_2}\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|\right)\leq\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon\\
\Rightarrow&\mathbb{P}\left(E_N\text{ and }\|\nabla\hat{H}_{K_1\times K_2}(\Lambda^*_{K_1\times K_2})\|<\frac{\bar{b}(\epsilon)}{8\eta_2}\left\|\Lambda-\Lambda^*_{K_1\times K_2}\right\|\right)>1-\epsilon\ . \qquad (33)
\end{aligned}$$

With (31) and (33), we can obtain that

$$\mathbb{P}\left\{\hat{G}_{K_1\times K_2}(\Lambda) - \hat{G}_{K_1\times K_2}(\Lambda^*_{K_1\times K_2}) < 0, \quad \forall\Lambda \in \partial\hat{\Upsilon}_{K_1\times K_2}(\epsilon)\right\} \geq 1-\epsilon \ .$$

Note that the event $\left\{\hat{G}_{K_1\times K_2}(\Lambda^*_{K_1\times K_2}) > \hat{G}_{K_1\times K_2}(\Lambda),\ \forall\Lambda \in \partial\hat{\Upsilon}_{K_1\times K_2}(\epsilon)\right\}$ implies that there exists a local maximizer in the interior of $\hat{\Upsilon}_{K_1\times K_2}(\epsilon)$. Since $\hat{G}_{K_1\times K_2}(\cdot)$ is strictly concave and $\hat{\Lambda}_{K_1\times K_2}$ is the unique global maximizer of $\hat{G}_{K_1\times K_2}$, then we get

$$\mathbb{P}\left(\hat{\Lambda}_{K_1\times K_2} \in \hat{\Upsilon}_{K_1\times K_2}(\epsilon)\right) > 1-\epsilon \ , \tag{34}$$

i.e. $\left\|\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\| = O_p\left(\sqrt{\frac{K}{N}}\right)$.

■

3.3 Corollary 3.3

The next corollary states that $\hat{\pi}_K(t,\boldsymbol{x})$ is arbitrarily close to $\pi^*_K(t,\boldsymbol{x})$.

Corollary 3.3 *Under Assumption 1.2-1.4, we have*

$$\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} |\hat{\pi}_K(t,\boldsymbol{x}) - \pi^*_K(t,\boldsymbol{x})|^2 = O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right),$$

and

$$\int_{\mathcal{T}\times\mathcal{X}} |\hat{\pi}_K(t,\boldsymbol{x}) - \pi^*_K(t,\boldsymbol{x})|^2 dF_{T,X}(t,\boldsymbol{x}) = O_p\left(\frac{K}{N}\right),$$

and

$$\frac{1}{N}\sum_{i=1}^{N} |\hat{\pi}_K(T_i,\boldsymbol{X}_i) - \pi^*_K(T_i,\boldsymbol{X}_i)|^2 = O_p\left(\frac{K}{N}\right).$$

Proof. From the proof of Lemma 3.2, we know the facts $\mathbb{P}\left(\hat{\Lambda}_{K_1\times K_2} \in \hat{\Upsilon}_{K_1\times K_2}(\epsilon)\right) > 1-\epsilon$ and (28). Then for any element $\tilde{\Lambda}_{K_1\times K_2}$ lying on the line joining $\hat{\Lambda}_{K_1\times K_2}$ and $\Lambda^*_{K_1\times K_2}$, we can have that $\mathbb{P}(u_{K_1}(t)^\top\tilde{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x}) \in \Gamma_2(\epsilon)$ for all $(t,\boldsymbol{x}) \in \mathcal{T}\times\mathcal{X}) \geq 1-\epsilon$, which implies

$$\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} |\rho''(u_{K_1}(t)\tilde{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x}))| = O_p(1). \tag{35}$$

Using Mean Value Theorem, Lemma 3.1, and (35), we can obtain that

$$\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} |\hat{\pi}_K(t,\boldsymbol{x}) - \pi^*_K(t,\boldsymbol{x})|$$

$$
\begin{aligned}
&= \sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} |\rho'\left(u_{K_1}(t)\hat{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right) - \rho'\left(u_{K_1}(t)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)| \\
&\leq \sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} |\rho''(u_{K_1}(t)\tilde{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x}))| \sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} \left|u_{K_1}(t)\hat{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x}) - u_{K_1}(t)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right| \\
&\leq O_p(1)\cdot\|\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\|\cdot\sup_{t\in\mathcal{T}}\|u_{K_1}(t)\|\cdot\sup_{\boldsymbol{x}\in\mathcal{X}}\|v_{K_2}(\boldsymbol{x})\| \\
&\leq O_p(1)\cdot O_p\left(\sqrt{\frac{K}{N}}\right)\zeta_1(K_1)\cdot\zeta_2(K_2) = O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right).
\end{aligned}
$$

Note that by Mean Value Theorem and (35), we can deduce that

$$
\begin{aligned}
&\int_{\mathcal{T}\times\mathcal{X}} |\hat{\pi}_K(t,\boldsymbol{x}) - \pi^*_K(t,\boldsymbol{x})|^2 dF_{T,X}(t,\boldsymbol{x}) \\
\leq & \sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} |\rho''(u_{K_1}(t)\tilde{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x}))|^2 \int_{\mathcal{T}\times\mathcal{X}} \left|u_{K_1}(t)\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{x})\right|^2 dF_{T,X}(t,\boldsymbol{x}) \\
\leq & O_p(1)\cdot\int_{\mathcal{T}\times\mathcal{X}} \left|u_{K_1}(t)\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{x})\right|^2 dF_{T,X}(t,\boldsymbol{x}).
\end{aligned}
$$

We estimate $\int_{\mathcal{T}\times\mathcal{X}} \left|u_{K_1}(t)\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{x})\right|^2 dF_{T,X}(t,\boldsymbol{x})$. Note that $\mathbb{E}[u_{K_1}(T)[u_{K_1}(T)^\top] = I_{K_1\times K_1}$, $\mathbb{E}[v_{K_2}(\boldsymbol{X})v_{K_2}(\boldsymbol{X})^\top] = I_{K_2\times K_2}$, (13) and Assumption 1.2, we can deduce that

$$
\begin{aligned}
&\int_{\mathcal{T}\times\mathcal{X}} \left|u_{K_1}(t)\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{x})\right|^2 dF_{T,X}(t,\boldsymbol{x}) \\
\leq & \int_{\mathcal{T}\times\mathcal{X}} u^\top_{K_1}(t)\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{x})v_{K_2}(\boldsymbol{x})^\top\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}^\top u_{K_1}(t)dF_{T,X}(t,\boldsymbol{x}) \\
= & \int_{\mathcal{T}\times\mathcal{X}} \frac{1}{\pi_0(t,\boldsymbol{x})}\pi_0(t,\boldsymbol{x})u^\top_{K_1}(t)\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{x})v_{K_2}(\boldsymbol{x})^\top\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}^\top u_{K_1}(t)dF_{T,X}(t,\boldsymbol{x}) \\
\leq & \frac{1}{\eta_1}\int_{\mathcal{T}\times\mathcal{X}} \pi_0(t,\boldsymbol{x})\cdot u^\top_{K_1}(t)\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{x})v_{K_2}(\boldsymbol{x})^\top\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}^\top u_{K_1}(t)dF_{T,X}(t,\boldsymbol{x}) \\
= & \frac{1}{\eta_1}\int_{\mathcal{T}} u^\top_{K_1}(t)\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}\left(\int_{\mathcal{X}} v_{K_2}(\boldsymbol{x})v_{K_2}(\boldsymbol{x})^\top dF_X(x)\right)\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}^\top u_{K_1}(t)dF_T(t) \\
= & \frac{1}{\eta_1}\int_{\mathcal{T}} u^\top_{K_1}(t)\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}^\top u_{K_1}(t)dF_T(t) \\
= & \frac{1}{\eta_1}\mathrm{tr}\left(\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}^\top\int_{\mathcal{T}} u_{K_1}(t)u^\top_{K_1}(t)dF_T(t)\right) \\
= & \frac{1}{\eta_1}\mathrm{tr}\left(\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\}^\top\right) \\
= & \frac{1}{\eta_1}\cdot\left\|\hat{\Lambda}_{K_2\times K_2} - \Lambda^*_{K_1\times K_2}\right\|^2 = O_p\left(\frac{K}{N}\right). \qquad (36)
\end{aligned}
$$

Then we obtain

$$\int_{\mathcal{T}\times\mathcal{X}} |\hat{\pi}_K(t,\boldsymbol{x}) - \pi_K^*(t,\boldsymbol{x})|^2 dF_{T,X}(t,\boldsymbol{x}) = O_p\left(\frac{K}{N}\right).$$

Similar to (23), we have

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N \left|u_{K_1}^\top(T_i)\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda_{K_1\times K_2}^*\right\} v_{K_2}(\boldsymbol{X}_i)\right|^2 - \int_{\mathcal{T}\times\mathcal{X}} \left|u_{K_1}(t)\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda_{K_1\times K_2}^*\right\} v_{K_2}(\boldsymbol{x})\right|^2 dF_{T,X}(t,\boldsymbol{x}) \\
=&O_p\left(\frac{\zeta(K)}{\sqrt{N}}\cdot\|\hat{\Lambda}_{K_1\times K_2} - \Lambda_{K_1\times K_2}^*\|^2\right) = O_p\left(\frac{\zeta(K)}{\sqrt{N}}\cdot\frac{K}{N}\right) = o_p\left(\frac{K}{N}\right). \quad (37)
\end{aligned}$$

where the last equality holds in light of Assumption 1.4. Hence, with (36) and (37), we have

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N |\hat{\pi}_K(T_i,\boldsymbol{X}_i) - \pi_K^*(T_i,\boldsymbol{X}_i)|^2 \\
\leq & \sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} |\rho''(u_{K_1}(t)\tilde{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x}))|^2 \cdot \frac{1}{N}\sum_{i=1}^N \left|u_{K_1}(T_i)\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda_{K_1\times K_2}^*\right\} v_{K_2}(\boldsymbol{X}_i)\right|^2 \\
\leq & O_p(1)\cdot\int_{\mathcal{T}\times\mathcal{X}} \left|u_{K_1}(t)\left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda_{K_1\times K_2}^*\right\} v_{K_2}(\boldsymbol{x})\right|^2 dF_{T,X}(t,\boldsymbol{x}) + o_p\left(\frac{K}{N}\right) \\
\leq & O_p\left(\frac{K}{N}\right) + o_p\left(\frac{K}{N}\right) = O_p\left(\frac{K}{N}\right).
\end{aligned}$$

■

4 Efficient Estimation

4.1 Proof of Theorem 4

Because $\widehat{\boldsymbol{\beta}}$ (resp. $\boldsymbol{\beta}^*$) is a unique minimizer of $N^{-1}\sum_{i=1}^N \widehat{\pi}_K(T_i,\boldsymbol{X}_i)L(Y_i - g(T_i;\boldsymbol{\beta}))$ (resp. $\mathbb{E}\left[\pi_0(T,\boldsymbol{X})L(Y - g(T;\boldsymbol{\beta}))\right]$), from the theory of M-estimation (van der Vaart, 1998, Theorem 5.7), if the following condition holds:

$$\sup_{\boldsymbol{\beta}\in\Theta}\left|\frac{1}{N}\sum_{i=1}^N \widehat{\pi}_K(T_i,\boldsymbol{X}_i)L(Y_i - g(T_i;\boldsymbol{\beta})) - \mathbb{E}\left[\pi_0(T,\boldsymbol{X})L(Y - g(T;\boldsymbol{\beta}))\right]\right| \xrightarrow{p} 0.$$

then $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}^*$. Note that

$$\sup_{\boldsymbol{\beta}\in\Theta}\left|\frac{1}{N}\sum_{i=1}^N \widehat{\pi}_K(T_i,\boldsymbol{X}_i)L(Y_i - g(T_i;\boldsymbol{\beta})) - \mathbb{E}\left[\pi_0(T,\boldsymbol{X})L(Y - g(T;\boldsymbol{\beta}))\right]\right|$$

$$\leq \sup_{\boldsymbol{\beta}\in\Theta}\left|\frac{1}{N}\sum_{i=1}^{N}\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}\,L\left(Y_i-g(T_i;\boldsymbol{\beta})\right)\right| \tag{38}$$

$$+\sup_{\boldsymbol{\beta}\in\Theta}\left|\frac{1}{N}\sum_{i=1}^{N}\pi_0(T_i,\boldsymbol{X}_i)L\left(Y_i-g(T_i;\boldsymbol{\beta})\right)-\mathbb{E}\left[\pi_0(T,\boldsymbol{X})L\left(Y-g(T;\boldsymbol{\beta})\right)\right]\right|. \tag{39}$$

We first show (38) is of $o_P(1)$. By Theorem 3, $\widehat{\pi}_K(\cdot)\xrightarrow{L^2(F_N)}\pi_0(\cdot)$, using Cauchy-Scharwz' inequality and Assumption 1.5, we have

$$\begin{aligned}|(38)|\leq&\left\{\frac{1}{N}\sum_{i=1}^{N}\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}^2\right\}^{1/2}\cdot\sup_{\boldsymbol{\beta}\in\Theta}\left\{\frac{1}{N}\sum_{i=1}^{N}L\left(Y_i-g(T_i;\boldsymbol{\beta})\right)^2\right\}^{1/2}\\ \leq& o_P(1)\cdot\left\{\sup_{\boldsymbol{\beta}\in\Theta}\mathbb{E}\left[L\left(Y-g(T;\boldsymbol{\beta})\right)^2\right]+o_P(1)\right\}^{1/2}=o_P(1)\quad\text{(by Assumption 5)}\end{aligned}$$

To show (39) is of $o_P(1)$, by (Newey and McFadden, 1994, Lemma 2.4), it is sufficient to require the following conditions holds:

1. Θ is compact;
2. $L(Y-g(T;\boldsymbol{\beta}))$ is continuous in $\boldsymbol{\beta}$;
3. $\mathbb{E}\left[\sup_{\boldsymbol{\beta}\in\Theta}|L(Y-g(T;\boldsymbol{\beta}))|\right]<\infty$.

which are the imposed Assumption 1.5.

4.2 Proof of Theorem 5

The proposed estimator $\hat{\boldsymbol{\beta}}$ is a special case of Chen, Linton, and Van Keilegom (2003), where the authors establish the consistency and asymptotic normality of a class of semiparametric optimization estimators under that the criterion function does not satisfy standard smoothness conditions. The asymptotic distribution of the proposed estimator can be derived by applying Theorem 2 of Chen, Linton, and Van Keilegom (2003).

Using their notation, we denote

$$\begin{aligned}M_N(\boldsymbol{\beta},\pi(\cdot)):=&\frac{1}{N}\sum_{i=1}^{N}\pi(T_i,\boldsymbol{X}_i)L'\left(Y_i-g(T_i;\boldsymbol{\beta})\right)m(T_i;\boldsymbol{\beta}),\\ M(\boldsymbol{\beta},\pi(\cdot)):=&\mathbb{E}[M_N(\boldsymbol{\beta},\pi(\cdot))]=\mathbb{E}\left[\pi(T,\boldsymbol{X})L'\left(Y-g(T;\boldsymbol{\beta})\right)m(T;\boldsymbol{\beta})\right].\end{aligned}$$

The ordinary derivative $\Gamma_1(\boldsymbol{\beta}, \pi(\cdot))$ in $\boldsymbol{\beta}$ of $M(\boldsymbol{\beta}, \pi(\cdot))$ is

$$\begin{aligned}\Gamma_1(\boldsymbol{\beta}, \pi(\cdot))(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}) =: &\lim_{\tau \to 0} \frac{M(\boldsymbol{\beta} + \tau(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}), \pi(\cdot)) - M(\boldsymbol{\beta}, \pi(\cdot))}{\tau} \\ =&\nabla_{\boldsymbol{\beta}} \mathbb{E}\left[\pi(T, \boldsymbol{X}) L'\left(Y - g(T; \boldsymbol{\beta})\right) m(T; \boldsymbol{\beta})\right],\end{aligned}$$

and the functional derivative $\Gamma_2(\boldsymbol{\beta}, \pi_0(\cdot))[\pi(\cdot) - \pi_0(\cdot)]$ of $M(\boldsymbol{\beta}, \pi_0(\cdot))$ along the direction $\pi(\cdot) - \pi_0(\cdot)$ is

$$\begin{aligned}\Gamma_2(\boldsymbol{\beta}, \pi_0(\cdot))[\pi(\cdot) - \pi_0(\cdot)] :=& \lim_{\tau \to 0} \frac{M(\boldsymbol{\beta}, \pi_0(\cdot) + \tau(\pi(\cdot) - \pi_0(\cdot))) - M(\boldsymbol{\beta}, \pi_0(\cdot))}{\tau} \\ =&\mathbb{E}\left[(\pi(T, \boldsymbol{X}) - \pi_0(T, \boldsymbol{X})) L'\left(Y - g(T; \boldsymbol{\beta})\right) m(T; \boldsymbol{\beta})\right].\end{aligned}$$

In order to apply Theorem 2 of Chen, Linton, and Van Keilegom (2003), we need to verify their Conditions (2.1)-(2.6) hold. Conditions (2.1)-(2.5) of Chen, Linton, and Van Keilegom (2003) can be easily verified by using following facts:

- Theorem 4 ensures $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\| \xrightarrow{p} 0$;
- Assumption 1.6 (iv) implies $\|M_N(\hat{\boldsymbol{\beta}}, \hat{\pi}_K(\cdot))\| = \left\| N^{-1} \sum_{i=1}^N \hat{\pi}_K(T_i, \boldsymbol{X}_i) m(T_i; \hat{\boldsymbol{\beta}}) L'\{Y_i - g(T_i; \hat{\boldsymbol{\beta}})\} \right\| = o_P(1/\sqrt{N})$;
- Assumption 1.8 implies $K = o_p(N^{1/2})$ and $K^{-\alpha} = o_p(N^{-1/2})$, then by Theorem 2 we have $\int_{\mathcal{T} \times \mathcal{X}} |\hat{\pi}_K(t, \boldsymbol{x}) - \pi_0(t, \boldsymbol{x})|^2 dF_{T,X}(t, \boldsymbol{x}) = O_p\left(K^{-\alpha}\right) + O_p\left(\sqrt{K/N}\right) = o_P(N^{-1/2}) + o_P(N^{-1/4}) \leq o_p(N^{-1/4})$.

The most important step toward the application of Theorem 2 of Chen, Linton, and Van Keilegom (2003) is to check their Condition (2.6) holds, which states that there exits some finite matrix V_1 such that

$$\sqrt{N}\left\{M_N(\boldsymbol{\beta}^*, \pi_0(\cdot)) + \Gamma_2(\boldsymbol{\beta}^*, \pi_0(\cdot))[\hat{\pi}_K(\cdot) - \pi_0(\cdot)]\right\} \xrightarrow{d} N(0, V_1). \tag{40}$$

If Conditions (2.1)-(2.6) hold, Theorem 2 of Chen, Linton, and Van Keilegom (2003) ensures that

$$\sqrt{N}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right) \xrightarrow{d} \mathcal{N}(0, \Omega),$$

where $\Omega := \Gamma_1(\boldsymbol{\beta}^*, \pi_0(\cdot))^{-1} V_1 (\Gamma_1(\boldsymbol{\beta}^*, \pi_0(\cdot))^{-1})^\top = H_0^{-1} V_1 (H_0^{-1})^\top$. However, Chen, Linton, and Van Keilegom (2003) do not give the expression of V_1 and the verification of (40) is difficult which is also admitted by the authors themselves (see the first paragraph in Section 3.3 of Chen, Linton, and Van Keilegom (2003)). In Section 4.3, we prove (40) holds and give

$$V_1 = \mathbb{E}[\psi(Y,T,\boldsymbol{X};\boldsymbol{\beta}^*)\psi(Y,T,\boldsymbol{X};\boldsymbol{\beta}^*)^\top].$$

Therefore, we can have $\Omega = V_{eff}$ which justifies Theorem 5.

4.3 Proof of (40)

Before proving (40), we prepare some preliminary notation and results that will be used later. Since $\hat{\Lambda}_{K_1\times K_2}$ is a unique maximizer of the concave function $\hat{G}_{K_1\times K_2}$, then

$$\frac{1}{N}\sum_{i=1}^N \rho'\left(u_{K_1}(T_i)^\top \hat{\Lambda}_{K_1\times K_2} v_{K_2}(\boldsymbol{X}_i)\right) u_{K_1}(T_i) v_{K_2}(\boldsymbol{X}_i)^\top - \frac{1}{N^2}\sum_{i=1}^N\sum_{l=1}^N u_{K_1}(T_l) v_{K_2}(\boldsymbol{X}_i)^\top = 0.$$

Using Mean Value Theorem, we can have

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N \rho'\left(u_{K_1}(T_i)^\top \Lambda^*_{K_1\times K_2} v_{K_2}(\boldsymbol{X}_i)\right) u_{K_1}(T_i) v_{K_2}(\boldsymbol{X}_i)^\top \\
&+\frac{1}{N}\sum_{i=1}^N \rho''\left(u_{K_1}(T_i)^\top \tilde{\Lambda}_{K_1\times K_2} v_{K_2}(\boldsymbol{X}_i)\right) u_{K_1}(T_i) u_{K_1}(T_i)^\top \left\{\hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}\right\} v_{K_2}(\boldsymbol{X}_i) v_{K_2}(\boldsymbol{X}_i)^\top \\
=&\frac{1}{N^2}\sum_{i=1}^N\sum_{l=1}^N u_{K_1}(T_l) v_{K_2}(\boldsymbol{X}_i)^\top \ , \qquad (41)
\end{aligned}$$

where $\tilde{\Lambda}_{K_1\times K_2}$ lies on the line joining from $\hat{\Lambda}_{K_1\times K_2}$ to $\Lambda^*_{K_1\times K_2}$. We define the following notation:

$$\hat{A}_{K_1\times K_2} := \hat{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}, \qquad (42)$$

$$\tilde{A}_{K_1\times K_2} := \tilde{\Lambda}_{K_1\times K_2} - \Lambda^*_{K_1\times K_2}, \qquad (43)$$

and

$$\begin{aligned}
&A^*_{K_1\times K_2} := \nabla \hat{G}_{K_1\times K_2}\left(\Lambda^*_{K_1\times K_2}\right) \\
=&\frac{1}{N}\sum_{i=1}^N \rho'\left(u_{K_1}(T_i)^\top \Lambda^*_{K_1\times K_2} v_{K_2}(\boldsymbol{X}_i)\right) u_{K_1}(T_i) v_{K_2}(\boldsymbol{X}_i)^\top - \left(\frac{1}{N}\sum_{l=1}^N u_{K_1}(T_l)\right)\left(\frac{1}{N}\sum_{i=1}^N v_{K_2}(\boldsymbol{X}_i)^\top\right). \qquad (44)
\end{aligned}$$

In light of (27) we have

$$\left\|A^*_{K_1\times K_2}\right\| = O_p\left(\sqrt{\frac{K}{N}}\right).$$

From (41), $A^*_{K_1\times K_2}$ can also be written as

$$A^*_{K_1\times K_2} = -\frac{1}{N}\sum_{i=1}^{N}\rho''\left(u_{K_1}(T_i)^\top\tilde{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)\right)u_{K_1}(T_i)u_{K_1}(T_i)^\top\left\{\hat{\Lambda}_{K_1\times K_2}-\Lambda^*_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{X}_i)v_{K_2}(\boldsymbol{X}_i)^\top. \tag{45}$$

We now start to (40). We decompose $\sqrt{N}\{M_N(\boldsymbol{\beta}^*,\pi_0(\cdot))+\Gamma_2(\boldsymbol{\beta}^*,\pi_0(\cdot))[\hat{\pi}_K(\cdot)-\pi_0(\cdot)]\}$ as follows:

$$\begin{aligned}
&\sqrt{N}\{M_N(\boldsymbol{\beta}^*,\pi_0(\cdot))+\Gamma_2(\boldsymbol{\beta}^*,\pi_0(\cdot))[\hat{\pi}_K(\cdot)-\pi_0(\cdot)]\}\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\pi_0(T_i,\boldsymbol{X}_i)L'\{Y_i-g(T_i;\boldsymbol{\beta}^*)\}m(T_i;\boldsymbol{\beta}^*)+\int_{\mathcal{T}}\int_{\mathcal{X}}(\hat{\pi}_K(t,\boldsymbol{x})-\pi_0(t,\boldsymbol{x}))\varepsilon(\boldsymbol{x},t;\boldsymbol{\beta}^*)m(t;\boldsymbol{\beta}^*)dF_{X,T}(\boldsymbol{x},t)\right\}\\
=&\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)(\pi^*_K(t,\boldsymbol{x})-\pi_0(t,\boldsymbol{x}))dF_{X,T}(\boldsymbol{x},t)\\
&+\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}(\hat{\pi}_K(t,\boldsymbol{x})-\pi^*_K(t,\boldsymbol{x}))\varepsilon(\boldsymbol{x},t;\boldsymbol{\beta}^*)m(t;\boldsymbol{\beta}^*)dF_{X,T}(\boldsymbol{x},t)\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\pi_0(T_i,\boldsymbol{X}_i)L'\{Y_i-g(T_i;\boldsymbol{\beta}^*)\}m(T_i;\boldsymbol{\beta}^*)\\
=&\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)(\pi^*_K(t,\boldsymbol{x})-\pi_0(t,\boldsymbol{x}))dF_{X,T}(\boldsymbol{x},t) &(46)\\
&+\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}(\hat{\pi}_K(t,\boldsymbol{x})-\pi^*_K(t,\boldsymbol{x}))\varepsilon(\boldsymbol{x},t;\boldsymbol{\beta}^*)m(t;\boldsymbol{\beta}^*)dF_{X,T}(\boldsymbol{x},t) &(47)\\
&\quad-\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho''\left(u^\top_{K_1}(t)\tilde{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)u^\top_{K_1}(t)\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})m(t;\boldsymbol{\beta}^*)dF_{X,T}(\boldsymbol{x},t)\\
&+\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho''\left(u^\top_{K_1}(t)\tilde{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)u^\top_{K_1}(t)\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})m(t;\boldsymbol{\beta}^*)dF_{X,T}(\boldsymbol{x},t) &(48)\\
&\quad-\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho''\left(u^\top_{K_1}(t)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)u^\top_{K_1}(t)A^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})m(t;\boldsymbol{\beta}^*)dF_{X,T}(\boldsymbol{x},t)\\
&+\sqrt{N}\int_{\mathcal{X}}\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho''\left(u^\top_{K_1}(t)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)u^\top_{K_1}(t)A^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})m(t;\boldsymbol{\beta}^*)dF_{X,T}(\boldsymbol{x},t) &(49)\\
&\quad+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big\{\pi_0(T_i,\boldsymbol{X}_i)m(T_i;\boldsymbol{\beta}^*)\varepsilon(T_i,\boldsymbol{X}_i;\boldsymbol{\beta}^*)-\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)|\boldsymbol{X}=\boldsymbol{X}_i\right]\\
&\qquad\qquad-\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)|T=T_i\right]\Big\}
\end{aligned}$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big\{\pi_0(T_i,\boldsymbol{X}_i)L'\left\{Y_i-g\left(T_i;\boldsymbol{\beta}^*\right)\right\}m(T_i;\boldsymbol{\beta}^*)-\pi_0(T_i,\boldsymbol{X}_i)m(T_i;\boldsymbol{\beta}^*)\varepsilon(T_i,\boldsymbol{X}_i;\boldsymbol{\beta}^*) \tag{50}$$
$$+\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)|\boldsymbol{X}=\boldsymbol{X}_i\right]+\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)|T=T_i\right]\Big\},$$

where $\hat{A}_{K_1\times K_2}$ and $A^*_{K_1\times K_2}$ are defined in (42) and (45). We show that the terms (46)-(49) are all of $o_p(1)$, while the term (50) is asymptotically normal.

For term (46)**:** By Lemma 3.1 and Assumption 1.4, we can deduce that

$$\begin{aligned}
&\left\|\sqrt{N}\cdot\mathbb{E}\left[m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\left(\pi_K^*(T,\boldsymbol{X})-\pi_0(T,\boldsymbol{X})\right)\right]\right\| \\
\leq&\sqrt{N}\sup_{t\in\mathcal{T}}\|m(t;\boldsymbol{\beta}^*)\|\cdot\mathbb{E}[|\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)|^2]^{\frac{1}{2}}\cdot\mathbb{E}\left[|\pi_K^*(T,\boldsymbol{X})-\pi_0(T,\boldsymbol{X})|^2\right]^{\frac{1}{2}}=O\left(\sqrt{N}K^{-\alpha}\right)\ .
\end{aligned}$$

For term (47)**:** By Mean Value Theorem and the definition of $\hat{A}_{K_1\times K_2}$ in (42), the term (47) is exactly equal to zero.

For term (48)**:** We can telescope (48) as follows:

$$\begin{aligned}
&\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho''\left(u_{K_1}^\top(t)\tilde{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)u_{K_1}^\top(t)\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t)\\
&-\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho''\left(u_{K_1}^\top(t)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)u_{K_1}^\top(t)A^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t)\\
=&\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\left\{\rho''\left(u_{K_1}^\top(t)\tilde{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)-\rho''\left(u_{K_1}^\top(t)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)\right\}\\
&\qquad\qquad\times u_{K_1}^\top(t)\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t) &(51)\\
&+\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho''\left(u_{K_1}^\top(t)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)u_{K_1}^\top(t)\\
&\qquad\qquad\times\left\{\hat{A}_{K_1\times K_2}-A^*_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t). &(52)
\end{aligned}$$

For the term (51), by Mean Value Theorem,

$$(51)=\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho'''\left(\xi_3(t,\boldsymbol{x})\right)\left\{u_{K_1}^\top(t)\tilde{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right\}\left\{u_{K_1}^\top(t)\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right\}dF_{X,T}(\boldsymbol{x},t).$$

Since $\xi_3(t,\boldsymbol{x})$ lies between $u_{K_1}(t)^\top\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})$ and $u_{K_1}(t)^\top\tilde{\Lambda}^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})$, which implies $\xi_3(t,\boldsymbol{x})$ lies between $u_{K_1}(t)^\top\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})$ and $u_{K_1}(t)^\top\hat{\Lambda}^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})$. Then in light of (28) and (34), we have $\mathbb{P}\left(\xi_3(t,\boldsymbol{x}) \in \Gamma_2(\epsilon),\ \forall (t,\boldsymbol{x}) \in \mathcal{T}\times\mathcal{X}\right) > 1-\epsilon$, therefore,

$$\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} |\rho'''\left(\xi_3(t,\boldsymbol{x})\right)| = O_p(1)\ . \tag{53}$$

With (36), (53), the fact $\|\tilde{A}_{K_1\times K_2}\| \leq \|\hat{A}_{K_1\times K_2}\|$, Lemma 3.2, and Assumption 1.4, we can derive that

$$\begin{aligned}
\|(51)\| \leq & \sqrt{N} \sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} |\rho'''\left(\xi_3(t,\boldsymbol{x})\right)| \sup_{t\in\mathcal{T}} \|m(t;\boldsymbol{\beta}^*)\| \cdot \sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} |\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)| \\
& \cdot \int_{\mathcal{T}}\int_{\mathcal{X}} \left| u_{K_1}(t)^\top \tilde{A}_{K_1\times K_2} v_{K_2}(\boldsymbol{x})\right| \cdot \left| u_{K_1}^\top(t) \hat{A}_{K_1\times K_2} v_{K_2}(\boldsymbol{x})\right| dF_{X,T}(\boldsymbol{x},t) \\
\leq & \sqrt{N}\cdot O_p(1)\cdot O(1)\cdot O(1)\cdot \left\{\int_{\mathcal{T}}\int_{\mathcal{X}} \left| u_{K_1}(t)^\top \tilde{A}_{K_1\times K_2} v_{K_2}(\boldsymbol{x})\right|^2 dF_{X,T}(\boldsymbol{x},t)\right\}^{\frac{1}{2}} \\
& \cdot \left\{\int_{\mathcal{T}}\int_{\mathcal{X}} \left| u_{K_1}^\top(t) \hat{A}_{K_1\times K_2} v_{K_2}(\boldsymbol{x})\right|^2 dF_{X,T}(\boldsymbol{x},t)\right\}^{\frac{1}{2}} \\
= & \sqrt{N}\cdot O_p(1)\cdot O(1)\cdot O(1)\cdot O_p\left(\sqrt{\frac{K}{N}}\right)\cdot O_p\left(\sqrt{\frac{K}{N}}\right) = O_p\left(\sqrt{\frac{K^2}{N}}\right) \qquad \text{(by ((36)))}.
\end{aligned} \tag{54}$$

For the term (52), we first compute the probability order of $\|A^*_{K_1\times K_2} - \hat{A}_{K_1\times K_2}\|$. Using (45), the fact $\rho''(v) = -\rho'(v)$ and Mean Value Theorem, we have

$$\begin{aligned}
& A^*_{K_1\times K_2} - \hat{A}_{K_1\times K_2} \\
= & -\frac{1}{N}\sum_{i=1}^N \rho''\left(u_{K_1}^\top(T_i)\Lambda^*_{K_1\times K_2} v_{K_2}(\boldsymbol{X}_i)\right) u_{K_1}(T_i) u_{K_1}(T_i)^\top \hat{A}_{K_1\times K_2} v_{K_2}(\boldsymbol{X}_i) v_{K_2}^\top(\boldsymbol{X}_i) \\
& -\frac{1}{N}\sum_{i=1}^N \rho'''\left(\xi_3(T_i,\boldsymbol{X}_i)\right)\left\{u_{K_1}(T)^\top \tilde{A}_{K_1\times K_2} v_{K_2}(\boldsymbol{X}_i)\right\} u_{K_1}(T_i) u_{K_1}(T_i)^\top \hat{A}_{K_1\times K_2} v_{K_2}(\boldsymbol{X}_i) v_{K_2}^\top(\boldsymbol{X}_i) \\
& - \hat{A}_{K_1\times K_2} \\
= & \frac{1}{N}\sum_{i=1}^N \left\{\rho'\left(u_{K_1}^\top(T_i)\Lambda^*_{K_1\times K_2} v_{K_2}(\boldsymbol{X}_i)\right) u_{K_1}(T_i) u_{K_1}(T_i)^\top \hat{A}_{K_1\times K_2} v_{K_2}(\boldsymbol{X}_i) v_{K_2}^\top(\boldsymbol{X}_i) - \hat{A}_{K_1\times K_2}\right\} && (55) \\
& -\frac{1}{N}\sum_{i=1}^N \rho'''\left(\xi_3(T_i,\boldsymbol{X}_i)\right)\left\{u_{K_1}(T_i)^\top \tilde{A}_{K_1\times K_2} v_{K_2}(\boldsymbol{X}_i)\right\} u_{K_1}(T_i) u_{K_1}(T_i)^\top \hat{A}_{K_1\times K_2} v_{K_2}(\boldsymbol{X}_i) v_{K_2}^\top(\boldsymbol{X}_i). && (56)
\end{aligned}$$

For the term (55), by (13) we can write $\hat{A}_{K_1\times K_2}$ as

$$\hat{A}_{K_1\times K_2} = \mathbb{E}_{T,\boldsymbol{X}}\left[\pi_0(T,\boldsymbol{X}) u_{K_1}(T) u_{K_1}(T)^\top \hat{A}_{K_1\times K_2} v_{K_2}(\boldsymbol{X}) v_{K_2}^\top(\boldsymbol{X})\right],$$

where $\mathbb{E}_{T,\boldsymbol{X}}[\cdot]$ denotes taking expectation with respect to $(T, \boldsymbol{X})$. We telescope (55) as follows:

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\rho'\left(u_{K_1}^{\top}(T_i)\Lambda_{K_1\times K_2}^{*}v_{K_2}(\boldsymbol{X}_i)\right)u_{K_1}(T_i)u_{K_1}(T_i)^{\top}\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)v_{K_2}^{\top}(\boldsymbol{X}_i)-\hat{A}_{K_1\times K_2}\right\}$$

$$=\frac{1}{N}\sum_{i=1}^{N}\left\{\left\{\rho'\left(u_{K_1}^{\top}(T_i)\Lambda_{K_1\times K_2}^{*}v_{K_2}(\boldsymbol{X}_i)\right)-\pi_0(T_i,\boldsymbol{X}_i)\right\}u_{K_1}(T_i)u_{K_1}(T_i)^{\top}\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)v_{K_2}^{\top}(\boldsymbol{X}_i)\right\} \tag{57}$$

$$-\frac{1}{N}\sum_{i=1}^{N}\left\{\pi_0(T_i,\boldsymbol{X}_i)u_{K_1}(T_i)u_{K_1}(T_i)^{\top}\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)v_{K_2}^{\top}(\boldsymbol{X}_i)\right.$$
$$\left.-\mathbb{E}_{T,\boldsymbol{X}}\left[\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^{\top}\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{X})v_{K_2}^{\top}(\boldsymbol{X})\right]\right\}. \tag{58}$$

For the term (57), by Lemmas 3.1 and 3.2 and (36), we have that

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\left\{\rho'\left(u_{K_1}^{\top}(T_i)\Lambda_{K_1\times K_2}^{*}v_{K_2}(\boldsymbol{X}_i)\right)-\pi_0(T_i,\boldsymbol{X}_i)\right\}u_{K_1}(T_i)u_{K_1}(T_i)^{\top}\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)v_{K_2}^{\top}(\boldsymbol{X}_i)\right\|$$
$$\leq\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|\rho'\left(u_{K_1}^{\top}(T_i)\Lambda_{K_1\times K_2}^{*}v_{K_2}(\boldsymbol{X}_i)\right)-\pi_0(T_i,\boldsymbol{X}_i)\right|^2\|u_{K_1}(T_i)\|^2\|v_{K_2}(\boldsymbol{X}_i)\|^2}$$
$$\cdot\sqrt{\frac{1}{N}\sum_{i=1}^{N}|u_{K_1}(T_i)^{\top}\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)|^2}$$
$$\leq O(\zeta(K)K^{-\alpha})\cdot\left[\int_{\mathcal{T}\times\mathcal{X}}|u_{K_1}(t)\widehat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})|^2dF_{T,X}(t,\boldsymbol{x})+\|\widehat{A}_{K_1\times K_2}\|^2\cdot O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right)\right]^{1/2}$$
$$\leq O(\zeta(K)K^{-\alpha})\cdot O_p(\|\widehat{A}_{K_1\times K_2}\|)$$
$$=O(\zeta(K)K^{-\alpha})\cdot O_p\left(\sqrt{\frac{K}{N}}\right)=O_p\left(N^{-\frac{1}{2}}\zeta(K)\cdot K^{\frac{1}{2}-\alpha}\right).$$

For the term (58), define the linear map $\mathcal{J}(\cdot):\mathbb{R}^{K_1\times K_2}\to\mathbb{R}$ by

$$\mathcal{J}(M):=\frac{1}{N}\sum_{i=1}^{N}\left\{\pi_0(T_i,\boldsymbol{X}_i)u_{K_1}(T_i)u_{K_1}(T_i)^{\top}Mv_{K_2}(\boldsymbol{X}_i)v_{K_2}^{\top}(\boldsymbol{X}_i)-\mathbb{E}_{T,\boldsymbol{X}}\left[\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^{\top}Mv_{K_2}(\boldsymbol{X})v_{K_2}^{\top}(\boldsymbol{X})\right]\right\},$$

then (58) $=\mathcal{J}(\hat{A}_{K_1\times K_2})$. For any fixed $M\in\mathbb{R}^{K_1\times K_2}$, by (13) and $M=\mathbb{E}[\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^{\top}M$ $\cdot v_{K_2}(\boldsymbol{X})v_{K_2}^{\top}(\boldsymbol{X})]$, then we have

$$\mathbb{E}\left[\mathcal{J}(M)^2\right]$$

$$
\begin{aligned}
=&\frac{1}{N}\cdot\mathbb{E}\left[\left\|\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^\top M v_{K_2}(\boldsymbol{X})v_{K_2}^\top(\boldsymbol{X})-\mathbb{E}\left[\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^\top M v_{K_2}(\boldsymbol{X})v_{K_2}^\top(\boldsymbol{X})\right]\right\|^2\right]\\
\leq&\frac{1}{N}\cdot\mathbb{E}\left[\left\|\pi_0(T,\boldsymbol{X})u_{K_1}(T)u_{K_1}(T)^\top M v_{K_2}(\boldsymbol{X})v_{K_2}^\top(\boldsymbol{X})\right\|^2\right]\\
\leq&\frac{1}{N}\cdot\eta_2\cdot\mathbb{E}\left[\pi_0(T,\boldsymbol{X})\cdot\|u_{K_1}(T)\|^4\|v_{K_2}(\boldsymbol{X})\|^4\right]\cdot\|M\|^2\\
=&\frac{1}{N}\cdot\eta_2\cdot\mathbb{E}[\|u_{K_1}(T)\|^4]\cdot\mathbb{E}[\|v_{K_2}(\boldsymbol{X})\|^4]\cdot\|M\|^2\\
\leq&\frac{1}{N}\cdot\eta_2\cdot\zeta_1(K)^2\cdot\zeta_2(K)^2\cdot\mathbb{E}[\|u_{K_1}(T)\|^2]\cdot\mathbb{E}[\|v_{K_2}(\boldsymbol{X})\|^2]\cdot\|M\|^2\\
=&\|M\|^2\cdot O\left(\zeta(K)^2\frac{K}{N}\right).
\end{aligned}
$$

Using Chebyshev's inequality we have

$$
|\mathcal{J}(M)|=\|M\|O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right)\ ,
$$

then in light of Lemma 3.2,

$$
(58)=\mathcal{J}(\hat{A}_{K_1\times K_2})=\|\hat{A}_{K_1\times K_2}\|O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right)=O_p\left(\zeta(K)\frac{K}{N}\right)\ .
$$

Therefore,

$$
(55)=(57)+(58)=O_p\left(N^{-\frac{1}{2}}\zeta(K)\cdot K^{\frac{1}{2}-\alpha}\right)+O_p\left(\zeta(K)\frac{K}{N}\right)\ .
$$

For the term (56), in light of (53) and Lemma 3.2, we can deduce that

$$
\begin{aligned}
&\left\|\frac{1}{N}\sum_{i=1}^N\rho'''\left(\xi_3(T_i,\boldsymbol{X}_i)\right)\left\{u_{K_1}(T_i)^\top\tilde{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)\right\}\left\{u_{K_1}(T_i)u_{K_1}(T_i)^\top\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)v_{K_2}^\top(\boldsymbol{X}_i)\right\}\right\|\\
\leq&\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}\left|\rho'''\left(\xi_3(t,\boldsymbol{x})\right)\right|\cdot\zeta(K)\cdot\frac{1}{N}\sum_{i=1}^N\left|u_{K_1}(T_i)^\top\tilde{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)\right|\cdot\left|u_{K_1}(T_i)^\top\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)\right|\\
\leq&\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}\left|\rho'''\left(\xi_3(t,\boldsymbol{x})\right)\right|\cdot\zeta(K)\cdot\sqrt{\frac{1}{N}\sum_{i=1}^N\left|u_{K_1}(T_i)^\top\tilde{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)\right|^2}\cdot\sqrt{\frac{1}{N}\sum_{i=1}^N\left|u_{K_1}(T_i)^\top\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)\right|^2}\\
\leq&O(1)\cdot\zeta(K)\cdot O_p(\|\tilde{A}_{K_1\times K_2}\|)\cdot O_p(\|\hat{A}_{K_1\times K_2}\|)\leq O_p(1)\cdot\zeta(K)\cdot O_p\left(\sqrt{\frac{K}{N}}\right)\cdot O_p\left(\sqrt{\frac{K}{N}}\right)\leq O_p\left(\zeta(K)\frac{K}{N}\right).
\end{aligned}
$$

Now, we can obtain

$$\|\hat{A}_{K_1\times K_2} - A^*_{K_1\times K_2}\| = (55) + (56) = O_p\left(N^{-\frac{1}{2}}\zeta(K)K^{\frac{1}{2}-\alpha}\right) + O_p\left(\zeta(K)\frac{K}{N}\right) + O_p\left(\zeta(K)\frac{K}{N}\right)$$
$$= O_p\left(N^{-\frac{1}{2}}\zeta(K)\cdot K^{\frac{1}{2}-\alpha}\right) + O_p\left(\zeta(K)\frac{K}{N}\right). \tag{59}$$

Using (59), Assumptions 1.7 and 1.4, for large enough N, we have

$$(52) = \left\|\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho''\left(u^\top_{K_1}(t)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)u^\top_{K_1}(t)\left\{\hat{A}_{K_1\times K_2} - A^*_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t)\right\|$$
$$\leq\sqrt{N}\sup_{t\in\mathcal{T}}\|m(t;\boldsymbol{\beta}^*)\|\sup_{\gamma\in\Gamma_1}|\rho''(\gamma)|\cdot\mathbb{E}\left[|\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)|^2\right]^{\frac{1}{2}}\cdot\left[\int_{\mathcal{T}\times\mathcal{X}}\left(u_{K_1}(t)\left\{\hat{A}_{K_1\times K_2} - A^*_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{x})\right)^2 dF_{T,X}(t,\boldsymbol{x})\right]^{\frac{1}{2}}$$
$$\leq\sqrt{N}\cdot O(1)\cdot O(1)\cdot O(1)\cdot O(1)\cdot O(\|\hat{A}_{K_1\times K_2} - A^*_{K_1\times K_2}\|)$$
$$\leq O_p\left(\zeta(K)\cdot K^{\frac{1}{2}-\alpha}\right) + O_p\left(\zeta(K)\frac{K}{\sqrt{N}}\right), \tag{60}$$

where the second inequality holds since by using the same argument of establishing (36), we have

$$\int_{\mathcal{T}\times\mathcal{X}}\left(u_{K_1}(t)\left\{\hat{A}_{K_1\times K_2} - A^*_{K_1\times K_2}\right\}v_{K_2}(\boldsymbol{x})\right)^2 dF_{T,X}(t,\boldsymbol{x}) = O(\|\hat{A}_{K_1\times K_2} - A^*_{K_1\times K_2}\|).$$

Therefore, by combining (54) and (60), we can obtain that

$$(48) = (51) + (52) = O_p\left(\sqrt{\frac{K^2}{N}}\right) + O_p\left(\zeta(K)\cdot K^{\frac{1}{2}-\alpha}\right) + O_p\left(\zeta(K)\frac{K}{\sqrt{N}}\right)$$
$$= O_p\left(\zeta(K)\cdot K^{\frac{1}{2}-\alpha}\right) + O_p\left(\zeta(K)\frac{K}{\sqrt{N}}\right).$$

***For term* (49):** By the definition of $A^*_{K_1\times K_2}$ in (44), we have

$$(49) = \frac{1}{\sqrt{N}}\sum_{i=1}^N\left\{\int_{\mathcal{T}}\int_{\mathcal{X}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho''\left(u^\top_{K_1}(t)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)u^\top_{K_1}(t)\right. \tag{61}$$
$$\left.\times\left\{u_{K_1}(T_i)\rho'\left(u_{K_1}(T_i)\Lambda^*_{K_1\times K_2}v_{K_2}(\boldsymbol{X}_i)\right)v^\top_{K_2}(\boldsymbol{X}_i)\right\}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t) + m(T_i;\boldsymbol{\beta}^*)\varepsilon(T_i,\boldsymbol{X}_i;\boldsymbol{\beta}^*)\pi_0(T_i,\boldsymbol{X}_i)\right\}$$
$$-\frac{1}{\sqrt{N}}\sum_{i=1}^N\left\{\int_{\mathcal{T}}\int_{\mathcal{X}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho''\left(u^\top_{K_1}(t)\Lambda^*_{K_1\times K_2}v^\top_{K_2}(\boldsymbol{x})\right)u^\top_{K_1}(t)\left(\frac{1}{N}\sum_{l=1}^N u_{K_1}(T_l)\right)\right. \tag{62}$$
$$\left.\times\left(\frac{1}{N}\sum_{j=1}^N v^\top_{K_2}(\boldsymbol{X}_j)\right)v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t) + \mathbb{E}\left[\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)|\boldsymbol{X}=\boldsymbol{X}_i\right]\right.$$

$$+\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)|T=T_i\right]\Bigg\}.$$

We shall show that both (61) and (62) are of $o_p(1)$. Noting $\rho''=-\rho'$, we can telescope (61) as follows:

$$\begin{aligned}
(61)=&\frac{1}{\sqrt{N}}\sum_{i=1}^N\Bigg\{\int_{\mathcal{T}}\int_{\mathcal{X}}m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho'\left(u_{K_1}^\top(t)\Lambda_{K_1\times K_2}^*v_{K_2}(\boldsymbol{x})\right)u_{K_1}^\top(t) &(63)\\
&\times\Big\{u_{K_1}(T_i)\Big[-\rho'\left(u_{K_1}(T_i)\Lambda_{K_1\times K_2}^*v_{K_2}(\boldsymbol{X}_i)\right)+\pi_0(T_i,\boldsymbol{X}_i)\Big]v_{K_2}^\top(\boldsymbol{X}_i)\Big\}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t)\Bigg\}\\
&-\frac{1}{\sqrt{N}}\sum_{i=1}^N\Bigg\{\int_{\mathcal{T}}\int_{\mathcal{X}}m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\left\{\rho'\left(u_{K_1}^\top(t)\Lambda_{K_1\times K_2}^*v_{K_2}(\boldsymbol{x})\right)-\pi_0(t,\boldsymbol{x})\right\}u_{K_1}^\top(t) &(64)\\
&\times\left\{u_{K_1}(T_i)\pi_0(T_i,\boldsymbol{X}_i)v_{K_2}^\top(\boldsymbol{X}_i)\right\}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t)\Bigg\}\\
&-\frac{1}{\sqrt{N}}\sum_{i=1}^N\Bigg\{\int_{\mathcal{T}}\int_{\mathcal{X}}m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\pi_0(t,\boldsymbol{x})u_{K_1}^\top(t)\left\{u_{K_1}(T_i)\pi_0(T_i,\boldsymbol{X}_i)v_{K_2}^\top(\boldsymbol{X}_i)\right\}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t)\\
&+m(T_i;\boldsymbol{\beta}^*)\varepsilon(T_i,\boldsymbol{X}_i;\boldsymbol{\beta}^*)\pi_0(T_i,\boldsymbol{X}_i)\Bigg\}\,. &(65)
\end{aligned}$$

We shall show that (63), (64) and (65) are all of $o_p(1)$. Note that second moment of (63) is

$$\begin{aligned}
&\mathbb{E}[|(63)|^2]=\mathbb{E}\Bigg[\Bigg|\int_{\mathcal{T}}\int_{\mathcal{X}}m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho'\left(u_{K_1}^\top(t)\Lambda_{K_1\times K_2}^*v_{K_2}(\boldsymbol{x})\right)u_{K_1}^\top(t)\\
&\qquad\times\Big\{u_{K_1}(T_i)\Big[-\rho'\left(u_{K_1}(T_i)\Lambda_{K_1\times K_2}^*v_{K_2}(\boldsymbol{X}_i)\right)+\pi_0(T_i,\boldsymbol{X}_i)\Big]v_{K_2}^\top(\boldsymbol{X}_i)\Big\}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t)\Bigg|^2\Bigg]\\
=&\mathbb{E}\Bigg[\Bigg|\int_{\mathcal{T}}\int_{\mathcal{X}}\pi_0(t,\boldsymbol{x})\cdot m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\left[\frac{\rho'\left(u_{K_1}^\top(t)\Lambda_{K_1\times K_2}^*v_{K_2}(\boldsymbol{x})\right)}{\pi_0(t,\boldsymbol{x})}\right]u_{K_1}^\top(t)\\
&\qquad\times\Big\{u_{K_1}(T_i)\Big[-\rho'\left(u_{K_1}(T_i)\Lambda_{K_1\times K_2}^*v_{K_2}(\boldsymbol{X}_i)\right)+\pi_0(T_i,\boldsymbol{X}_i)\Big]v_{K_2}^\top(\boldsymbol{X}_i)\Big\}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t)\Bigg|^2\Bigg]\\
\leq&\mathbb{E}\Bigg[\Bigg|\int_{\mathcal{T}}\int_{\mathcal{X}}\pi_0(t,\boldsymbol{x})\cdot m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\left[\frac{\rho'\left(u_{K_1}^\top(t)\Lambda_{K_1\times K_2}^*v_{K_2}(\boldsymbol{x})\right)}{\pi_0(t,\boldsymbol{x})}\right]u_{K_1}^\top(t)\left\{u_{K_1}(T_i)v_{K_2}^\top(\boldsymbol{X}_i)\right\}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t)\Bigg|^2\Bigg]\\
&\quad\times\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}\left\{-\rho'\left(u_{K_1}^\top(t)\Lambda_{K_1\times K_2}^*v_{K_2}(\boldsymbol{x})\right)+\pi_0(t,\boldsymbol{x})\right\}^2\\
=&\left\{\mathbb{E}\left[\left|m(T_i;\boldsymbol{\beta}^*)\varepsilon(T_i,\boldsymbol{X}_i;\boldsymbol{\beta}^*)\left[\frac{\rho'\left(u_{K_1}^\top(T_i)\Lambda_{K_1\times K_2}^*v_{K_2}(\boldsymbol{X}_i)\right)}{\pi_0(T_i,\boldsymbol{X}_i)}\right]\right|^2\right]+o(1)\right\}\times\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}\left\{-\pi_K^*(t,\boldsymbol{x})+\pi_0(t,\boldsymbol{x})\right\}^2\\
=&O(1)\cdot O(K^{-2\alpha}\zeta(K)^2)=O(K^{-2\alpha}\zeta(K)^2)\to0,\ \text{(by Assumption 1.4)}
\end{aligned}$$

where the third equality holds because

$$\int_{\mathcal{T}}\int_{\mathcal{X}}\pi_0(t,\boldsymbol{x})\cdot m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\left[\frac{\rho'\left(u_{K_1}^\top(t)\Lambda_{K_1\times K_2}^*v_{K_2}(\boldsymbol{x})\right)}{\pi_0(t,\boldsymbol{x})}\right]u_{K_1}^\top(t)\left\{u_{K_1}(T)v_{K_2}^\top(\boldsymbol{X})\right\}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t)$$

is the weighted L^2-projection of $m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\left[\frac{\rho'\left(u_{K_1}^\top(t)\Lambda_{K_1\times K_2}^* v_{K_2}(\boldsymbol{x})\right)}{\pi_0(t,\boldsymbol{x})}\right]$ on the space linearly spanned by $\{u_{K_1}(t), v_{K_2}(\boldsymbol{x})\}$ with the weighted measure $\pi_0(t,\boldsymbol{x})dF_{T,X}(t,\boldsymbol{x})$. Similarly, we can also show (64) and (65) are of $o_p(1)$. Therefore, (61) is of $o_p(1)$.

For the term (62), since $\rho''(v) = -\rho'(v)$ and the fact $\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\right] = 0$, we telescope it as follows:

$$
\begin{aligned}
(62) =& \sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho'\left(u_{K_1}^\top(t)\Lambda_{K_1\times K_2}^* v_{K_2}(\boldsymbol{x})\right) u_{K_1}^\top(t)\left(\frac{1}{N}\sum_{l=1}^N u_{K_1}(T_l) - \mathbb{E}\left[u_{K_1}(T)\right]\right) \\
& \times \left(\frac{1}{N}\sum_{j=1}^N v_{K_2}^\top(\boldsymbol{X}_j) - \mathbb{E}[v_{K_2}^\top(\boldsymbol{X})]\right) v_{K_2}(\boldsymbol{x}) dF_{X,T}(\boldsymbol{x},t) \qquad (66) \\
& + \frac{1}{\sqrt{N}}\sum_{i=1}^N \Big\{\int_{\mathcal{T}}\int_{\mathcal{X}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho'\left(u_{K_1}^\top(t)\Lambda_{K_1\times K_2}^* v_{K_2}(\boldsymbol{x})\right) u_{K_1}^\top(t)\mathbb{E}\left[u_{K_1}(T)\right] v_{K_2}^\top(\boldsymbol{X}_i) v_{K_2}(\boldsymbol{x}) dF_{X,T}(\boldsymbol{x},t) \\
& \qquad - \mathbb{E}\left[\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)|\boldsymbol{X}=\boldsymbol{X}_i\right]\Big\} \qquad (67) \\
& + \frac{1}{\sqrt{N}}\sum_{l=1}^N \Big\{\int_{\mathcal{T}}\int_{\mathcal{X}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho'\left(u_{K_1}^\top(t)\Lambda_{K_1\times K_2}^* v_{K_2}(\boldsymbol{x})\right) u_{K_1}^\top(t) u_{K_1}(T_l)\mathbb{E}[v_{K_2}^\top(\boldsymbol{X})] v_{K_2}(\boldsymbol{x}) dF_{X,T}(\boldsymbol{x},t) \\
& \qquad - \mathbb{E}\left[\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)|T=T_i\right]\Big\} \qquad (68) \\
& - \frac{1}{\sqrt{N}}\sum_{i=1}^N \Big\{\int_{\mathcal{T}}\int_{\mathcal{X}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\rho'\left(u_{K_1}^\top(t)\Lambda_{K_1\times K_2}^* v_{K_2}(\boldsymbol{x})\right) u_{K_1}^\top(t)\mathbb{E}\left[u_{K_1}(T)\right]\mathbb{E}[v_{K_2}^\top(\boldsymbol{X})] v_{K_2}(\boldsymbol{x}) dF_{X,T}(\boldsymbol{x},t) \\
& \qquad - \mathbb{E}\left[\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)\right]\Big\}. \qquad (69)
\end{aligned}
$$

For the term (66), since

$$
\begin{aligned}
\left\|\frac{1}{N}\sum_{l=1}^N u_{K_1}(T_l) - \mathbb{E}\left[u_{K_1}(T)\right]\right\| &= O_p\left(\sqrt{\frac{K_1}{N}}\right), \\
\left\|\frac{1}{N}\sum_{j=1}^N v_{K_2}(\boldsymbol{X}_j) - \mathbb{E}\left[v_{K_2}(\boldsymbol{X})\right]\right\| &= O_p\left(\sqrt{\frac{K_2}{N}}\right), \\
\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}\left|\rho'\left(u_{K_1}^\top(t)\Lambda_{K_1\times K_2}^* v_{K_2}(\boldsymbol{x})\right)\right| &= O(1),
\end{aligned}
$$

and by Assumptions 1.4, 1.6 and 1.7, we can deduce that

$$
(66) = \sqrt{N}\cdot O(\zeta(K))O_p\left(\sqrt{\frac{K_1}{N}}\right)O_p\left(\sqrt{\frac{K_2}{N}}\right) = O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right) = o_p(1).
$$

For the term (67), noting the fact that $\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m(T;\boldsymbol{\beta}^*)\varepsilon(T,\boldsymbol{X};\boldsymbol{\beta}^*)|\boldsymbol{X}\right] = \int_{\mathcal{T}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{X};\boldsymbol{\beta}^*)$ $dF_T(t)$, we can rewrite (67) as follows:

$$
\begin{aligned}
(67) =& \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left\{\int_{\mathcal{T}}\int_{\mathcal{X}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\frac{\pi_K^*(t,\boldsymbol{x})}{\pi_0(t,\boldsymbol{x})}u_{K_1}^\top(t)\mathbb{E}\left[u_{K_1}(T)\right] v_{K_2}^\top(\boldsymbol{X}_j)v_{K_2}(\boldsymbol{x})dF_X(\boldsymbol{x})dF_T(t)\right.\\
&\left. - \int_{\mathcal{T}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{X}_j;\boldsymbol{\beta}^*)dF_T(t)\right\}.
\end{aligned}
$$

By computing the second moment of (67), we can obtain that

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\int_{\mathcal{T}}\int_{\mathcal{X}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\frac{\pi_K^*(t,\boldsymbol{x})}{\pi_0(t,\boldsymbol{x})}u_{K_1}^\top(t)\mathbb{E}\left[u_{K_1}(T)\right] v_{K_2}^\top(\boldsymbol{X})v_{K_2}(\boldsymbol{x})dF_X(\boldsymbol{x})dF_T(t) - \int_{\mathcal{T}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{X};\boldsymbol{\beta}^*)dF_T(t)\right\|^2\right]\\
\leq&\mathbb{E}\left[\left\|\int_{\mathcal{T}}\int_{\mathcal{X}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\frac{\pi_K^*(t,\boldsymbol{x})}{\pi_0(t,\boldsymbol{x})}u_{K_1}^\top(t)u_{K_1}(T^*)v_{K_2}^\top(\boldsymbol{X}^*)v_{K_2}(\boldsymbol{x})dF_X(\boldsymbol{x})dF_T(t) - m(T^*;\boldsymbol{\beta}^*)\varepsilon(T^*,\boldsymbol{X}^*;\boldsymbol{\beta}^*)\right\|^2\right]\\
\leq&2\cdot\mathbb{E}\left[\left\|\int_{\mathcal{T}}\int_{\mathcal{X}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)u_{K_1}^\top(t)u_{K_1}(T^*)v_{K_2}^\top(\boldsymbol{X}^*)v_{K_2}(\boldsymbol{x})dF_X(\boldsymbol{x})dF_T(t) - m(T^*;\boldsymbol{\beta}^*)\varepsilon(T^*,\boldsymbol{X}^*;\boldsymbol{\beta}^*)\right\|^2\right]\\
&+2\cdot\mathbb{E}\left[\left\|\int_{\mathcal{T}}\int_{\mathcal{X}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)\frac{\pi_K^*(t,\boldsymbol{x})-\pi_0(t,\boldsymbol{x})}{\pi_0(t,\boldsymbol{x})}u_{K_1}^\top(t)u_{K_1}(T^*)v_{K_2}^\top(\boldsymbol{X}^*)v_{K_2}(\boldsymbol{x})dF_X(\boldsymbol{x})dF_T(t)\right\|^2\right]\to 0,
\end{aligned}
$$

where $T^* \sim F_T$, $\boldsymbol{X}^* \sim F_X$, and T^* is independent of $\boldsymbol{X}^*$; the first inequality holds by Jensen's inequality; the last convergence result follows from Lemma 3.1 and the fact that

$$
\int_{\mathcal{T}}\int_{\mathcal{X}} m(t;\boldsymbol{\beta}^*)\varepsilon(t,\boldsymbol{x};\boldsymbol{\beta}^*)u_{K_1}^\top(t)u_{K_1}(T^*)v_{K_2}^\top(\boldsymbol{X}^*)v_{K_2}(\boldsymbol{x})dF_X(\boldsymbol{x})dF_T(t)
$$

is the L^2-projection of $m(T^*;\boldsymbol{\beta}^*)\varepsilon(T^*,\boldsymbol{X}^*;\boldsymbol{\beta}^*)$ on the space spanned by $\{u_{K_1}(T^*), v_{K_2}(\boldsymbol{X}^*)\}$. Thus (67) is of $o_p(1)$ by Chebyshev's inequality. Similar argument can be applied to show that both (68) and (69) are of $o_p(1)$. Therefore, we can have that

$$
|(62)| \leq |(66)| + |(67)| + |(68)| = o_p(1)\ .
$$

Then, we can obtain that

$$
|(49)| \leq |(61)| + |(62)| = o_p(1)\ .
$$

Summing up all orders (46)-(49) and using Assumption 1.8, we have

$$
\begin{aligned}
&(46)+(47)+(48)+(49)\\
=&O(\sqrt{N}K^{-\alpha}) + 0 + \left\{O_p\left(\zeta(K)\cdot K^{\frac{1}{2}-\alpha}\right) + O_p\left(\zeta(K)\frac{K}{\sqrt{N}}\right)\right\} + o_p(1) = o_p(1).
\end{aligned}
$$

5 Some Extensions

5.1 Proof of Theorem 7

(Consistency). Let

$$\widehat{\gamma}=\left[\sum_{i=1}^{N}u_{K_1}(T_i)u_{K_1}(T_i)^\top\right]^{-1}\left[\sum_{i=1}^{N}u_{K_1}(T_i)\hat{\pi}_K(T_i,\boldsymbol{X}_i)Y_i\right]$$

then $\widehat{\theta}_t=\hat{\gamma}^\top u_{K_1}(t)$. By assumption, there exists $\gamma^*\in\mathbb{R}^{K_1}$ such that

$$\sup_{t\in\mathcal{T}}\left|\theta_t-(\gamma^*)^\top u_{K_1}(t)\right|=O(K_1^{-\tilde{\alpha}}). \tag{70}$$

We first claim that

$$\|\hat{\gamma}-\gamma^*\|=O_p\left(\zeta(K)\left\{\sqrt{\frac{K}{N}}+K^{-\alpha}\right\}+K_1^{-\tilde{\alpha}}\right)\ , \tag{71}$$

whose proof will be established later. Using (70) and (71), we can deduce that

$$\begin{aligned}
&\int_{\mathcal{T}}\left[\widehat{\theta}_t-\theta_t\right]^2 dF_T(t)\\
=&\int_{\mathcal{T}}[\hat{\gamma}^\top u_{K_1}(t)-(\gamma^*)^\top u_{K_1}(t)+(\gamma^*)^\top u_{K_1}(t)-\theta_t]^2 dF_T(t)\\
\leq&2(\hat{\gamma}-\gamma^*)^\top\left[\int_{\mathcal{T}}u_{K_1}(t)u_{K_1}(t)^\top dF_T(t)\right](\hat{\gamma}-\gamma^*)+2\int_{\mathcal{T}}[(\gamma^*)^\top u_{K_1}(t)-\theta_t]^2 dF_T(t)\\
\leq&2\|\hat{\gamma}-\gamma^*\|^2\cdot\lambda_{\max}\left(\mathbb{E}[u_{K_1}(T)u_{K_1}(T)^\top]\right)+2\sup_{t\in\mathcal{T}}|(\gamma^*)^\top u_{K_1}(t)-\theta_t|^2\\
=&O_p\left(\zeta(K)^2\left\{\frac{K}{N}+K^{-2\alpha}\right\}+K_1^{-2\tilde{\alpha}}\right),
\end{aligned}$$

and

$$\begin{aligned}
\sup_{t\in\mathcal{T}}|\hat{\theta}_t-\theta_t|=&\sup_{t\in\mathcal{T}}\left|\hat{\gamma}^\top u_{K_1}(t)-(\gamma^*)^\top u_{K_1}(t)+(\gamma^*)^\top u_{K_1}(t)-\theta_t\right|\\
\leq&\sup_{t\in\mathcal{T}}\|u_{K_1}(t)\|\cdot\|\hat{\gamma}-\gamma^*\|+\sup_{t\in\mathcal{T}}|(\gamma^*)^\top u_{K_1}(t)-\theta_t|
\end{aligned}$$

$$
\begin{aligned}
&\leq O_p\left[\zeta_1(K_1)\left(\zeta(K)\left\{\sqrt{\frac{K}{N}}+K^{-\alpha}\right\}+K_1^{-\tilde{\alpha}}\right)\right]+O(K_1^{-\tilde{\alpha}})\\
&=O_p\left[\zeta_1(K_1)\left(\zeta(K)\left\{\sqrt{\frac{K}{N}}+K^{-\alpha}\right\}+K_1^{-\tilde{\alpha}}\right)\right].
\end{aligned}
$$

Finally, we come back to prove (71). Note that

$$
\begin{aligned}
&\hat{\gamma}-\gamma^* = \left[\sum_{i=1}^N u_{K_1}(T_i)u_{K_1}(T_i)^\top\right]^{-1}\left[\sum_{i=1}^N \hat{\pi}_K(T_i,\boldsymbol{X}_i)u_{K_1}(T_i)Y_i\right]-\gamma^*\\
&=\left[\sum_{i=1}^N u_{K_1}(T_i)u_{K_1}(T_i)^\top\right]^{-1}\left[\sum_{i=1}^N u_{K_1}(T_i)\left\{\hat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\right\}Y_i\right]\\
&\quad+\left[\sum_{i=1}^N u_{K_1}(T_i)u_{K_1}(T_i)^\top\right]^{-1}\left[\sum_{i=1}^N u_{K_1}(T_i)\{\pi_0(T_i,\boldsymbol{X}_i)Y_i-\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)Y_i|T_i]\}\right]\\
&\quad+\left[\sum_{i=1}^N u_{K_1}(T_i)u_{K_1}(T_i)^\top\right]^{-1}\left[\sum_{i=1}^N u_{K_1}(T_i)\left\{\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)Y_i|T_i]-(\gamma^*)^\top u_{K_1}(T_i)\right\}\right]\\
&\equiv A_{1N}+A_{2N}+A_{3N}.
\end{aligned}
$$

We first compute the probability order of A_{1N}. We use the following notation:

$$
\begin{aligned}
&\hat{H}_N := \Big(\left\{\hat{\pi}_K(T_1,X_1)-\pi_0(T_1,X_1)\right\}Y_1,...,\left\{\hat{\pi}_K(T_N,X_N)-\pi_0(T_N,X_N)\right\}Y_N\Big)^\top,\\
&U_{N\times K_1} := (u_{K_1}(T_1),...,u_{K_1}(T_N))^\top,\\
&\hat{\Phi}_{K_1\times K_1} := \frac{1}{N}\sum_{i=1}^N u_{K_1}(T)u_{K_1}^\top(T).
\end{aligned}
$$

Then we can obtain that

$$
\begin{aligned}
\|A_{1N}\|^2 &= \left\|\left[\sum_{i=1}^N u_{K_1}(T_i)u_{K_1}(T_i)^\top\right]^{-1}\left[\sum_{i=1}^N u_{K_1}(T_i)\left\{\hat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\right\}Y_i\right]\right\|^2\\
&=N^{-2}\mathrm{tr}\left(\hat{\Phi}^{-1}_{K_1\times K_1}U^\top_{N\times K_1}\hat{H}_N\hat{H}_N^\top U_{N\times K_1}\hat{\Phi}^{-1}_{K_1\times K_1}\right)\\
&=N^{-2}\mathrm{tr}\left(U^\top_{N\times K_1}\hat{H}_N\hat{H}_N^\top U_{N\times K_1}\hat{\Phi}^{-1}_{K_1\times K_1}\hat{\Phi}^{-1}_{K_1\times K_1}\right)\\
&=N^{-2}\mathrm{tr}\left(\hat{\Phi}^{-1/2}_{K_1\times K_1}U^\top_{N\times K_1}\hat{H}_N\hat{H}_N^\top U_{N\times K_1}\hat{\Phi}^{-1/2}_{K_1\times K_1}\hat{\Phi}^{-1}_{K_1\times K_1}\right)\\
&\leq\lambda_{\max}(\hat{\Phi}^{-1}_{K_1\times K_1})N^{-2}\mathrm{tr}\left(\hat{\Phi}^{-1/2}_{K_1\times K_1}U^\top_{N\times K_1}\hat{H}_N\hat{H}_N^\top U_{N\times K_1}\hat{\Phi}^{-1/2}_{K_1\times K_1}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \lambda_{\max}(\hat{\Phi}_{K_1\times K_1}^{-1}) N^{-1} \mathrm{tr}\left(\hat{H}_N \hat{H}_N^\top U_{N\times K_1}(U_{N\times K_1}^\top U_{N\times K_1})^{-1} U_{N\times K_1}^\top\right) \\
&\leq [\lambda_{\min}(\hat{\Phi}_{K_1\times K_1})]^{-1} N^{-1} \|\hat{H}_N\|^2 \\
&= [\lambda_{\min}(\hat{\Phi}_{K_1\times K_1})]^{-1} \cdot \frac{1}{N}\sum_{i=1}^{N} \left\{\hat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\right\}^2 Y_i^2 \\
&\leq [\lambda_{\min}(\hat{\Phi}_{K_1\times K_1})]^{-1} \sup_{(t,x)\in\mathcal{T}\times\mathcal{X}} |\hat{\pi}_K(t,x) - \pi_0(t,x)|^2 \cdot \frac{1}{N}\sum_{i=1}^{N} Y_i^2 \\
&\leq O_p(1)\cdot O_p\left(\zeta(K)^2 K^{-2\alpha} + \frac{\zeta(K)^2 K}{N}\right)\cdot O_p(1) \\
&= O_p\left(\zeta(K)^2 K^{-2\alpha} + \frac{\zeta(K)^2 K}{N}\right), \qquad (72)
\end{aligned}
$$

where the first inequality follows from the fact that $tr(AB) \leq \lambda_{\max}(B) tr(A)$ for any symmetric matrix B and positive semidefinite matrix A, the second inequality follows from the same fact and the fact that $U_{N\times K_1}(U_{N\times K_1}^\top U_{N\times K_1})^{-1} U_{N\times K_1}^\top$ is a projection matrix with maximum eigenvalue 1, and the fourth inequality follows from the facts that $|\lambda_{\min}(\hat{\Phi}_{K_1\times K_1})|^{-1} = O_p(1)$, Lemma 3.1 and Corollary 3.3, and $N^{-1}\sum_{i=1}^{N} Y_i^2 = O_p(1)$.

Next, we compute the probability order of A_{2N}. Let

$$
\varepsilon_i := \pi_0(T_i, \boldsymbol{X}_i) Y_i - \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) Y_i | T_i] \text{ and } \mathcal{E}_N := (\varepsilon_1, \ldots, \varepsilon_N)^\top.
$$

We can deduce that

$$
\begin{aligned}
\|A_{2N}\|^2 &= \left\| \left[\sum_{i=1}^{N} u_{K_1}(T_i) u_{K_1}(T_i)^\top\right]^{-1} \left[\sum_{i=1}^{N} u_{K_1}(T_i)\varepsilon_i\right] \right\|^2 \\
&= N^{-2}\mathrm{tr}\left(\hat{\Phi}_{K_1\times K_1}^{-1} U_{N\times K_1}^\top \mathcal{E}_N \mathcal{E}_N^\top U_{N\times K_1} \hat{\Phi}_{K_1\times K_1}^{-1}\right) \\
&= N^{-2}\mathrm{tr}\left(U_{N\times K_1}^\top \mathcal{E}_N \mathcal{E}_N^\top U_{N\times K_1} \hat{\Phi}_{K_1\times K_1}^{-1} \hat{\Phi}_{K_1\times K_1}^{-1}\right) \\
&\leq [\lambda_{\min}(\hat{\Phi}_{K_1\times K_1})]^{-2} N^{-2} \|U_{N\times K_1}^\top \mathcal{E}_N\|^2 = O_p\left(\frac{K_1}{N}\right),
\end{aligned}
$$

where the last equality follows that $|\lambda_{\min}(\hat{\Phi}_{K_1\times K_1})|^{-1} = O_p(1)$ and $N^{-2}\|U_{N\times K_1}^\top \mathcal{E}_N\|^2 = O_p(K_1/N)$ by Markov's inequality.

We finally compute the probability order of A_{3N}. Let

$$R_N(\gamma^*) = \left(\left\{\mathbb{E}[\pi_0(T_1, X_1)Y_1|T_1] - (\gamma^*)^\top u_{K_1}(T_1)\right\}, \ldots, \left\{\mathbb{E}[\pi_0(T_N, X_N)Y_N|T_N] - (\gamma^*)^\top u_{K_1}(T_N)\right\} \right)^\top,$$

then

$$\begin{aligned}
\|A_{3N}\|^2 =& \left\| \left[\sum_{i=1}^N u_{K_1}(T_i) u_{K_1}(T_i)^\top \right]^{-1} \left[\sum_{i=1}^N u_{K_1}(T_i) \left\{ \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)Y_i|T_i] - (\gamma^*)^\top u_{K_1}(T_i) \right\} \right] \right\|^2 \\
=& N^{-2} \left\| \hat{\Phi}^{-1}_{K_1\times K_1} U^\top_{N\times K_1} R_N(\gamma^*) \right\|^2 \\
=& N^{-2} \mathrm{tr}\left(\hat{\Phi}^{-1}_{K_1\times K_1} U^\top_{N\times K_1} R_N(\gamma^*) R_N(\gamma^*)^\top U_{N\times K_1} \hat{\Phi}^{-1}_{K_1\times K_1} \right) \\
=& N^{-2} \mathrm{tr}\left(U^\top_{N\times K_1} R_N(\gamma^*) R_N(\gamma^*)^\top U_{N\times K_1} \hat{\Phi}^{-1}_{K_1\times K_1} \hat{\Phi}^{-1}_{K_1\times K_1} \right) \\
=& N^{-2} \mathrm{tr}\left(\hat{\Phi}^{-1/2}_{K_1\times K_1} U^\top_{N\times K_1} R_N(\gamma^*) R_N(\gamma^*)^\top U_{N\times K_1} \hat{\Phi}^{-1/2}_{K_1\times K_1} \hat{\Phi}^{-1}_{K_1\times K_1} \right) \\
\leq& \lambda_{\max}(\hat{\Phi}^{-1}_{K_1\times K_1}) N^{-2} \mathrm{tr}\left(\hat{\Phi}^{-1/2}_{K_1\times K_1} U^\top_{N\times K_1} R_N(\gamma^*) R_N(\gamma^*)^\top U_{N\times K_1} \hat{\Phi}^{-1/2}_{K_1\times K_1} \right) \\
=& \lambda_{\max}(\hat{\Phi}^{-1}_{K_1\times K_1}) N^{-1} \mathrm{tr}\left(R_N(\gamma^*) R_N(\gamma^*)^\top U_{N\times K_1} (U^\top_{N\times K_1} U_{N\times K_1})^{-1} U^\top_{N\times K_1} \right) \\
\leq& [\lambda_{\min}(\hat{\Phi}_{K_1\times K_1})]^{-1} N^{-1} \|R_N(\gamma^*)\|^2 \\
=& [\lambda_{\min}(\hat{\Phi}_{K_1\times K_1})]^{-1} \cdot \frac{1}{N} \sum_{i=1}^N \left\{ \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)Y_i|T_i] - (\gamma^*)^\top u_{K_1}(T_i) \right\}^2 = O_p(K_1^{-2\tilde{\alpha}}),
\end{aligned}$$

where the first 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, the second inequality follows from the same fact and the fact that $U_{N\times K_1}(U^\top_{N\times K_1} U_{N\times K_1})^{-1} U^\top_{N\times K_1}$ is a projection matrix with maximum eigenvalue 1, and the last equality follows from the fact that $|\lambda_{\min}(\hat{\Phi}_{K_1\times K_1})|^{-1} = O_p(1)$ and the fact that $\frac{1}{N}\sum_{i=1}^N \left\{ \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)Y_i|T_i] - (\gamma^*)^\top u_{K_1}(T_i) \right\}^2 \leq \sup_{t\in\mathcal{T}} |\mathbb{E}[\pi_0(T, X)Y|T = t] - (\gamma^*)^\top u_{K_1}(t)|^2 = O(K_1^{-2\tilde{\alpha}})$. Hence, we complete the proof of (71).

(Asymptotic Normality). We have the following decomposition for $\widehat{\theta}_t - \theta(t)$:

$$\begin{aligned}
&\hat{\theta}_t - \theta_t = u_{K_1}(t)^\top (\hat{\gamma} - \gamma^*) + [(\gamma^*)^\top u_{K_1}(t) - \theta_t] \\
=& u_{K_1}(t)^\top \left[\frac{1}{N} \sum_{i=1}^N u_{K_1}(T_i) u_{K_1}(T_i)^\top \right]^{-1} \left[\frac{1}{N} \sum_{i=1}^N u_{K_1}(T_i) \left\{ \hat{\pi}_K(T_i, \boldsymbol{X}_i)Y_i - \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i)Y_i|T_i\right] \right\} \right]
\end{aligned}$$

$$
\begin{aligned}
&+u_{K_1}(t)^\top\left[\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)u_{K_1}(T_i)^\top\right]^{-1}\left[\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)\cdot\left\{\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)Y_i|T_i\right]-(\gamma^*)^\top u_{K_1}(T_i)\right\}\right]\\
&\quad+\left[(\gamma^*)^\top u_{K_1}(t)-\theta_t\right]\\
&\equiv b_{1N}(t)+b_{2N}(t)+b_{3N}(t).
\end{aligned}
$$

We shall show that $b_{1N}(t)$ contributes to the asymptotic variance; and $b_{2N}(t)+b_{3N}(t)$ contributes to the asymptotic bias which is asymptotically negligible. Thus to complete the proof of asymptotic normality, it is sufficient to prove the following results:

(i) $V_t \geq c\|u_{K_1}(t)\|^2$ for some $c>0$;

(ii) $\sqrt{N}V_t^{-1/2}b_{1N}(t)\xrightarrow{d}N(0,1)$;

(iii) $\sqrt{N}V_t^{-1/2}b_{2N}(t)=o_p(1)$;

(iv) $\sqrt{N}V_t^{-1/2}b_{3N}(t)=o_p(1)$.

We first prove Result (i). By assumption, $\lambda_{\min}\left(\mathbb{E}\left[b_{K_1}(T,\boldsymbol{X},Y)b_{K_1}^\top(T,\boldsymbol{X},Y)\right]\right)\geq \underline{c}$, we have

$$
\begin{aligned}
V_t =&u_{K_1}^\top(t)\Phi_{K_1\times K_1}^{-1}\mathbb{E}\left[b_{K_1}(T,\boldsymbol{X},Y)b_{K_1}^\top(T,\boldsymbol{X},Y)\right]\Phi_{K_1\times K_1}^{-1}u_{K_1}(t)\\
\geq&\underline{c}\cdot u_{K_1}^\top(t)\Phi_{K_1\times K_1}^{-1}\Phi_{K_1\times K_1}^{-1}u_{K_1}(t)\\
\geq&\underline{c}\cdot\lambda_{\min}^2\left(\Phi_{K_1\times K_1}^{-1}\right)\|u_{K_1}(t)\|^2.
\end{aligned}
$$

For the claim (ii). Let

$$
\tilde{b}_{1N}(t)=u_{K_1}(t)^\top\left[\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)u_{K_1}(T_i)^\top\right]^{-1}\left[\frac{1}{N}\sum_{i=1}^N b_{K_1}(T_i,\boldsymbol{X}_i,Y_i)\right].
$$

Similar to the proof of (40), we can show

$$
\sqrt{N}V_t^{-1/2}\cdot(b_{1N}(t)-\tilde{b}_{1N}(t))=o_p(1)\ .
$$

Then

$$
\sqrt{N}V_t^{-1/2}b_{1N}(t)=\sqrt{N}V_t^{-1/2}\tilde{b}_{1N}(t)+o_p(1)
$$

$$
\begin{aligned}
&= \sqrt{N} V_t^{-1/2} u_{K_1}(t)^\top \hat{\Phi}_{K_1 \times K_1}^{-1} N^{-1} \sum_{i=1}^N b_{K_1}(T_i, \boldsymbol{X}_i, Y_i) \\
&= \sqrt{N} V_t^{-1/2} u_{K_1}(t)^\top \Phi_{K_1 \times K_1}^{-1} \cdot N^{-1} \sum_{i=1}^N b_{K_1}(T_i, \boldsymbol{X}_i, Y_i) \\
&\quad + \sqrt{N} V_t^{-1/2} u_{K_1}(t)^\top \left[\hat{\Phi}_{K_1 \times K_1}^{-1} - \Phi_{K_1 \times K_1}^{-1} \right] \cdot N^{-1} \sum_{i=1}^N b_{K_1}(T_i, \boldsymbol{X}_i, Y_i) \\
&\equiv b_{1N}^{(1)}(t) + b_{1N}^{(2)}(t).
\end{aligned} \tag{73}
$$

For $b_{1N}^{(1)}(t)$, we can simply apply the Liapounov CLT and show that $b_{1N}^{(1)}(t) \xrightarrow{d} N(0,1)$. For $b_{1N}^{(2)}(t)$, we can deduce that

$$
\begin{aligned}
|b_{1N,2}^{(2)}(t)|^2 \leq & \left\{ V_K^{-1} \|u_{K_1}(t)\|^2 \right\} \cdot \left\| \hat{\Phi}_{K_1 \times K_1}^{-1} - \Phi_{K_1 \times K_1}^{-1} \right\|^2 \cdot \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N b_{K_1}(T_i, \boldsymbol{X}_i, Y_i) \right\|^2 \\
\leq & O_p(1) \cdot O_P\left(\zeta_1(K_1)^2 \cdot \frac{K_1}{N} \right) \cdot O_P(K_1) = O_P\left(\zeta_1(K_1)^2 \cdot \frac{K_1^2}{N} \right) = o_P(1),
\end{aligned}
$$

where the second inequality by noting the following facts

$$
\begin{aligned}
& \left\| \hat{\Phi}_{K_1 \times K_1}^{-1} - \Phi_{K_1 \times K_1}^{-1} \right\|^2 \\
=& \operatorname{tr}\left(\left\{ \hat{\Phi}_{K_1 \times K_1}^{-1} - \Phi_{K_1 \times K_1}^{-1} \right\} \left\{ \hat{\Phi}_{K_1 \times K_1}^{-1} - \Phi_{K_1 \times K_1}^{-1} \right\} \right) \\
=& \operatorname{tr}\left(\hat{\Phi}_{K_1 \times K_1}^{-1} \left\{ \hat{\Phi}_{K_1 \times K_1} - \Phi_{K_1 \times K_1} \right\} \Phi_{K_1 \times K_1}^{-1} \Phi_{K_1 \times K_1}^{-1} \left\{ \hat{\Phi}_{K_1 \times K_1} - \Phi_{K_1 \times K_1} \right\} \hat{\Phi}_{K_1 \times K_1}^{-1} \right) \\
=& \operatorname{tr}\left(\left\{ \hat{\Phi}_{K_1 \times K_1} - \Phi_{K_1 \times K_1} \right\} \Phi_{K_1 \times K_1}^{-1} \Phi_{K_1 \times K_1}^{-1} \left\{ \hat{\Phi}_{K_1 \times K_1} - \Phi_{K_1 \times K_1} \right\} \hat{\Phi}_{K_1 \times K_1}^{-1} \hat{\Phi}_{K_1 \times K_1}^{-1} \right) \\
\leq & \lambda_{\min}\left(\hat{\Phi}_{K_1 \times K_1} \right)^{-2} \lambda_{\min}\left(\Phi_{K_1 \times K_1} \right)^{-2} \cdot \operatorname{tr}\left(\left\{ \hat{\Phi}_{K_1 \times K_1} - \Phi_{K_1 \times K_1} \right\} \left\{ \hat{\Phi}_{K_1 \times K_1} - \Phi_{K_1 \times K_1} \right\} \right) \\
\leq & O_p(1) \cdot O_p(1) \cdot O_P\left(\zeta_1(K_1)^2 \cdot \frac{K_1}{N} \right) = O_P\left(\zeta_1(K_1)^2 \cdot \frac{K_1}{N} \right),
\end{aligned}
$$

and

$$
\mathbb{E}\left[\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N b_{K_1}(T_i, \boldsymbol{X}_i, Y_i) \right\|^2 \right] = \mathbb{E}\left[\|b_{K_1}(T, \boldsymbol{X}, Y)\|^2 \right] = O(K_1).
$$

Thus (ii) holds.

For (iii), by Cauchy-Schwarz's inequality, we can obtain that

$$
\begin{aligned}
&\sqrt{N}V_t^{-1/2}|b_{2N}(t)| \\
=&N^{-1/2}V_t^{-1/2}\left|u_{K_1}(t)^\top\hat{\Phi}_{K_1\times K_1}^{-1}U_{N\times K_1}^\top R_N(\gamma^*)\right| \\
\leq&V_t^{-1/2}\left\{u_{K_1}(t)^\top\hat{\Phi}_{K_1\times K_1}^{-1}\left(N^{-1}U_{N\times K_1}^\top U_{N\times K_1}\right)\hat{\Phi}_{K_1\times K_1}^{-1}u_{K_1}(t)\right\}^{\frac{1}{2}}\left\{R_N(\gamma^*)^\top R_N(\gamma^*)\right\}^{\frac{1}{2}} \\
\leq&V_t^{-1/2}\left\{u_{K_1}(t)^\top\hat{\Phi}_{K_1\times K_1}^{-1}u_{K_1}(t)\right\}^{\frac{1}{2}}\left\{R_N(\gamma^*)^\top R_N(\gamma^*)\right\}^{\frac{1}{2}} \\
\leq&\{V_t^{-1/2}\|u_{K_1}(t)\|\}\cdot|\lambda_{\max}(\Phi_{K_1\times K_1}^{-1})|^{\frac{1}{2}}\cdot O(\sqrt{N}\cdot K_1^{-\tilde{\alpha}}) \\
=&O(1)\cdot O_p(1)\cdot o_p(1)=o_p(1)\ .
\end{aligned}
$$

Similarly, we can show show that $\sqrt{N}V_t^{-1/2}|b_{3N}(t)|=o_p(1)$. This completes the proof of the Theorem.

5.2 Proof of Theorem 9

Note that

$$
\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot\widehat{\theta}_{t_0,t_1|t_0}=\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot\widehat{\theta}_{t_1|t_0}-\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot\widehat{\theta}_{t_0|t_0}.
$$

Consider the term $\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot\widehat{\theta}_{t_0|t_0}$. Since $\widehat{\theta}_{t_0|t_0}$ is a nonparametric series estimator of $\theta_{t_0|t_0}$, by using a similar argument of proving Theorem 6 (see also Newey (1997)), we have

$$
\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot\widehat{\theta}_{t_0|t_0}=V_{t_1,t_0|t_0}^{-1/2}\cdot u_{K_1}(t_0)^\top\Phi_{K_1\times K_1}^{-1}\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^N b_{3,K_1}(T_i,Y_i)+o_P(1), \tag{74}
$$

where

$$
b_{3,K_1}(T_i,Y_i)=u_{K_1}(T_i)\{Y_i-\mathbb{E}[Y_i|T_i]\}.
$$

Consider the term $\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot\widehat{\theta}_{t_1|t_0}$. Let $\delta:=t_1-t_0$, and

$$
\widehat{\gamma}:=\left[\sum_{i=1}^N\frac{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)}{\widehat{\pi}_K(T_i-\delta,\boldsymbol{X}_i)}\cdot Y_i\cdot u_{K_1}^\top(T_i)\right]\left[\sum_{i=1}^N u_{K_1}(T_i)u_{K_1}^\top(T_i)\right]^{-1}.
$$

Then $\widehat{\theta}_{t_1|t_0}=\widehat{\gamma}^\top u_{K_1}(t_1)$. We have the following decomposition for $\widehat{\theta}_{t_1|t_0}-\theta_{t_1|t_0}$:

$$
\widehat{\theta}_{t_1|t_0}-\theta_{t_1|t_0}=u_{K_1}(t_1)^\top(\hat{\gamma}-\gamma^*)+[(\gamma^*)^\top u_{K_1}(t_1)-\theta_{t_1|t_0}]
$$

$$
\begin{aligned}
=& u_{K_1}(t_1)^\top \left[\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)u_{K_1}(T_i)^\top\right]^{-1}\left[\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)\left\{\frac{\hat{\pi}_K(T_i,\boldsymbol{X}_i)}{\hat{\pi}_K(T_i-\delta,\boldsymbol{X}_i)}Y_i - \frac{\hat{\pi}_K(T_i,\boldsymbol{X}_i)}{\pi_0(T_i-\delta,\boldsymbol{X}_i)}Y_i\right\}\right]\\
&+u_{K_1}(t_1)^\top \left[\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)u_{K_1}(T_i)^\top\right]^{-1}\left[\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)\left\{\frac{\hat{\pi}_K(T_i,\boldsymbol{X}_i)}{\pi_0(T_i-\delta,\boldsymbol{X}_i)}Y_i - \mathbb{E}\left[\frac{\pi_0(T_i,\boldsymbol{X}_i)}{\pi_0(T_i-\delta,\boldsymbol{X}_i)}Y_i\middle|T_i\right]\right\}\right]\\
&+u_{K_1}(t_1)^\top \left[\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)u_{K_1}(T_i)^\top\right]^{-1}\left[\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)\cdot\left\{\mathbb{E}\left[\frac{\pi_0(T_i,\boldsymbol{X}_i)}{\pi_0(T_i-\delta,\boldsymbol{X}_i)}Y_i\middle|T_i\right] - (\gamma^*)^\top u_{K_1}(T_i)\right\}\right]\\
&+\left[(\gamma^*)^\top u_{K_1}(t_1)-\theta_{t_1|t_0}\right]\\
\equiv& b_{1N}(t_1)+b_{2N}(t_1)+b_{3N}(t_1)+b_{4N}(t_1).
\end{aligned}
$$

Similar to the proof of Theorem 7 (pp 50, (iii) and (iv)), we can show

$$
\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}|b_{3N}(t_1)| \to 0 \text{ and } \sqrt{N}V_{t_1,t_0|t_0}^{-1/2}|b_{4N}(t_1)| \to 0. \tag{75}
$$

Consider $b_{1N}(t_1)$. Similar to (73), we can show that

$$
\begin{aligned}
&\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot b_{1N}(t_1)\\
=&-\sqrt{N}u_{K_1}(t_1)^\top\Phi_{K_1\times K_1}^{-1}\left[\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)\left\{\frac{\hat{\pi}_K(T_i-\delta,\boldsymbol{X}_i)-\pi_0(T_i-\delta,\boldsymbol{X}_i)}{\pi_0(T_i-\delta,\boldsymbol{X}_i)\hat{\pi}_K(T_i-\delta,\boldsymbol{X}_i)}\right\}\hat{\pi}_K(T_i,\boldsymbol{X}_i)Y_i\right]+o_P(1).
\end{aligned}
$$

Then we have

$$
\begin{aligned}
&\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot b_{1N}(t_1)\\
=&-\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot u_{K_1}(t_1)^\top\Phi_{K_1\times K_1}^{-1}\left[\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)\left\{\frac{\hat{\pi}_K(T_i-\delta,\boldsymbol{X}_i)-\pi_0(T_i-\delta,\boldsymbol{X}_i)}{\pi_0(T_i-\delta,\boldsymbol{X}_i)^2}\right\}\pi_0(T_i,\boldsymbol{X}_i)Y_i\right]\\
&-\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot u_{K_1}(t_1)^\top\Phi_{K_1\times K_1}^{-1}\left[\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)\left\{\frac{\hat{\pi}_K(T_i-\delta,\boldsymbol{X}_i)-\pi_0(T_i-\delta,\boldsymbol{X}_i)}{\pi_0(T_i-\delta,\boldsymbol{X}_i)}\right\}\right.\\
&\qquad\left.\times\left\{\frac{\hat{\pi}_K(T_i,\boldsymbol{X}_i)}{\hat{\pi}_K(T_i-\delta,\boldsymbol{X}_i)}-\frac{\pi_0(T_i,\boldsymbol{X}_i)}{\pi_0(T_i-\delta,\boldsymbol{X}_i)}\right\}Y_i\right]\\
=&-\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot u_{K_1}(t_1)^\top\Phi_{K_1\times K_1}^{-1}\left[\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)\left\{\frac{\hat{\pi}_K(T_i-\delta,\boldsymbol{X}_i)-\pi_0(T_i-\delta,\boldsymbol{X}_i)}{\pi_0(T_i-\delta,\boldsymbol{X}_i)^2}\right\}\pi_0(T_i,\boldsymbol{X}_i)Y_i\right]\\
&-\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot u_{K_1}(t_1)^\top\Phi_{K_1\times K_1}^{-1}\left[\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)\left\{\frac{\hat{\pi}_K(T_i-\delta,\boldsymbol{X}_i)-\pi_0(T_i-\delta,\boldsymbol{X}_i)}{\pi_0(T_i-\delta,\boldsymbol{X}_i)}\right\}\right.
\end{aligned}
$$

$$
\times \left\{ \frac{\hat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)}{\hat{\pi}_K(T_i - \delta, \boldsymbol{X}_i)} \right\} Y_i \Bigg]
$$

$$
+ \sqrt{N} V_{t_1,t_0|t_0}^{-1/2} \cdot u_{K_1}(t_1)^\top \Phi_{K_1\times K_1}^{-1} \left[\frac{1}{N} \sum_{i=1}^{N} u_{K_1}(T_i) \left\{ \frac{\hat{\pi}_K(T_i - \delta, \boldsymbol{X}_i) - \pi_0(T_i - \delta, \boldsymbol{X}_i)}{\pi_0(t_0, \boldsymbol{X}_i)} \right\} \right.
$$

$$
\left. \cdot \left\{ \frac{\hat{\pi}_K(T_i - \delta, \boldsymbol{X}_i) - \pi_0(T_i - \delta, \boldsymbol{X}_i)}{\pi_0(T_i - \delta, X_i) \hat{\pi}_K(T_i - \delta, \boldsymbol{X}_i)} \right\} \pi_0(T_i, \boldsymbol{X}_i) Y_i \right]
$$

$$
\equiv \sqrt{N} V_{t_1,t_0|t_0}^{-1/2} \cdot b_{1N}^{(1)}(t_1) + \sqrt{N} V_{t_1,t_0|t_0}^{-1/2} \cdot b_{1N}^{(2)}(t_1) + \sqrt{N} V_{t_1,t_0|t_0}^{-1/2} \cdot b_{1N}^{(3)}(t_1).
$$

Consider $\sqrt{N} V_{t_1,t_0|t_0}^{-1/2} \cdot b_{1N}^{(2)}(t_1)$. The conditions $\lambda_{\min}(\Sigma_{2K_1\times 2K_1}) > \underline{c} > 0$ and $\lambda_{\min}(\Phi_{K_1\times K_1}) > \underline{c} > 0$ imply $V_{t_1,t_0|t_0}^{-1} \geq c \cdot \|u_{K_1}(t_1)\|^2$ for some $c > 0$. Similar to (72), we can show that

$$
\sqrt{N} V_{t_1,t_0|t_0}^{-1/2} \left| b_{1N}^{(2)}(t_1) \right| \leq \sqrt{N} \left\{ V_{t_1,t_0|t_0}^{-1/2} \cdot \|u_{K_1}(t_1)\| \right\}
$$

$$
\cdot \left\| \Phi_{K_1\times K_1}^{-1} \left[\frac{1}{N} \sum_{i=1}^{N} u_{K_1}(T_i) \cdot \left\{ \frac{\hat{\pi}_K(T_i - \delta, \boldsymbol{X}_i) - \pi_0(T_i - \delta, \boldsymbol{X}_i)}{\pi_0(T_i - \delta, \boldsymbol{X}_i)} \right\} \left\{ \frac{\hat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)}{\hat{\pi}_K(T_i - \delta, \boldsymbol{X}_i)} \right\} Y_i \right] \right\|
$$

$$
\leq \sqrt{N} \cdot O_P(1) \cdot \left\{ \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{\hat{\pi}_K(T_i - \delta, \boldsymbol{X}_i) - \pi_0(T_i - \delta, \boldsymbol{X}_i)}{\pi_0(T_i - \delta, \boldsymbol{X}_i)} \right\}^2 \cdot \left\{ \frac{\hat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)}{\hat{\pi}_K(T_i - \delta, \boldsymbol{X}_i)} \right\}^2 Y_i^2 \right\}^{1/2}
$$

$$
\leq \sqrt{N} \cdot O_P(1) \cdot \sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} |\hat{\pi}_K(t,\boldsymbol{x}) - \pi_0(t,\boldsymbol{x})|^2 \cdot \left\{ \frac{1}{N} \sum_{i=1}^{N} Y_i^2 \right\}^{1/2}
$$

$$
\leq O_P\left(\sqrt{N} \cdot \zeta^2(K) \cdot \left\{ K^{-2\alpha} + \frac{K}{N} \right\} \right) = o_P(1). \tag{76}
$$

Similarly, we can also show

$$
\sqrt{N} \cdot V_{t_1,t_0|t_0}^{-1/2} \cdot \left| b_{1N}^{(3)}(t_1) \right| = o_P(1). \tag{77}
$$

We next consider $b_{1N}^{(1)}(t_1)$. We shall find the influence representation for $N^{-1/2} u_{K_1}(t_1) \Phi_{K_1\times K_1}^{-1} \cdot \sum_{i=1}^{N} u_{K_1}(T_i) \{\hat{\pi}_K(T_i - \delta, \boldsymbol{X}_i) - \pi_0(T_i - \delta, \boldsymbol{X}_i)\} \pi_0(T_i, \boldsymbol{X}_i) Y_i / \pi_0(t_0, \boldsymbol{X}_i)^2$. To achieve this goal, we consider the asymptotic behavior of $N^{-1/2} \sum_{i=1}^{N} \{\hat{\pi}_K(T_i - \delta, \boldsymbol{X}_i) \phi(T_i, \boldsymbol{X}_i, Y_i) - \mathbb{E}[\pi_0(T - \delta, \boldsymbol{X}) \phi(T, \boldsymbol{X}, Y)]\}$, where $\phi(T, \boldsymbol{X}, Y)$ denotes a general L^2 random variable. Define $\mu(t, \boldsymbol{x}) := \mathbb{E}[\phi(T, \boldsymbol{X}, Y) | T = t, \boldsymbol{X} = \boldsymbol{x}]$. Similar to the proof of (40) in Section 4.3, we have the following decomposition:

$$
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{\hat{\pi}_K(T_i - \delta, \boldsymbol{X}_i) \phi(T_i, \boldsymbol{X}_i, Y_i) - \mathbb{E}[\pi_0(T - \delta, \boldsymbol{X}) \phi(T, \boldsymbol{X}, Y)]\}
$$

$$
\begin{aligned}
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Bigg\{\left(\hat{\pi}_K(T_i-\delta,\boldsymbol{X}_i)-\pi_K^*(T_i-\delta,\boldsymbol{X}_i)\right)\phi(T_i,\boldsymbol{X}_i,Y_i) \\
&\quad-\int_{\mathcal{T}}\int_{\mathcal{X}}\left(\hat{\pi}_K(t-\delta,\boldsymbol{x})-\pi_K^*(t-\delta,\boldsymbol{x})\right)\mu(\boldsymbol{x},t)dF_{X,T}(\boldsymbol{x},t)\Bigg\}
\end{aligned} \tag{78}
$$

$$
\begin{aligned}
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Bigg\{\left(\pi_K^*(T_i-\delta,\boldsymbol{X}_i)-\pi_0(T_i-\delta,\boldsymbol{X}_i)\right)\phi(T_i,\boldsymbol{X}_i,Y_i) \\
&\qquad-\int_{\mathcal{T}}\int_{\mathcal{X}}\mu(t,\boldsymbol{x})\left(\pi_K^*(t-\delta,\boldsymbol{x})-\pi_0(t-\delta,\boldsymbol{x})\right)dF_{X,T}(\boldsymbol{x},t)\Bigg\}
\end{aligned} \tag{79}
$$

$$
+\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}\mu(t,\boldsymbol{x})\left(\pi_K^*(t-\delta,\boldsymbol{x})-\pi_0(t-\delta,\boldsymbol{x})\right)dF_{X,T}(\boldsymbol{x},t) \tag{80}
$$

$$
\begin{aligned}
&+\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}\left(\hat{\pi}_K(t-\delta,\boldsymbol{x})-\pi_K^*(t-\delta,\boldsymbol{x})\right)\mu(\boldsymbol{x},t)dF_{X,T}(\boldsymbol{x},t) \\
&-\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}\mu(t,\boldsymbol{x})\rho''\left(u_{K_1}^{\top}(t-\delta)\tilde{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)u_{K_1}^{\top}(t-\delta)\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t)
\end{aligned} \tag{81}
$$

$$
\begin{aligned}
&+\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}\mu(t,\boldsymbol{x})\rho''\left(u_{K_1}^{\top}(t-\delta)\tilde{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)u_{K_1}^{\top}(t-\delta)\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t) \\
&-\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}\mu(t,\boldsymbol{x})\rho''\left(u_{K_1}^{\top}(t-\delta)\Lambda_{K_1\times K_2}^*v_{K_2}(\boldsymbol{x})\right)u_{K_1}^{\top}(t-\delta)A_{K_1\times K_2}^*v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t)
\end{aligned} \tag{82}
$$

$$
\begin{aligned}
&+\sqrt{N}\int_{\mathcal{X}}\mu(t,\boldsymbol{x})\rho''\left(u_{K_1}^{\top}(t-\delta)\Lambda_{K_1\times K_2}^*v_{K_2}(\boldsymbol{x})\right)u_{K_1}^{\top}(t-\delta)A_{K_1\times K_2}^*v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t) \\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Bigg\{\pi_0(T_i,\boldsymbol{X}_i)\frac{f_{T|X}(T_i+\delta|\boldsymbol{X}_i)}{f_{T|X}(T_i|\boldsymbol{X}_i)}\mu(T_i+\delta,\boldsymbol{X}_i) \\
&\quad-\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\frac{f_{T|X}(T_i+\delta|\boldsymbol{X}_i)}{f_{T|X}(T_i|\boldsymbol{X}_i)}\mu(T_i+\delta,\boldsymbol{X}_i)\middle|\boldsymbol{X}_i\right] \\
&\quad-\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\frac{f_{T|X}(T_i+\delta|\boldsymbol{X}_i)}{f_{T|X}(T_i|\boldsymbol{X}_i)}\mu(T_i+\delta,\boldsymbol{X}_i)\middle|T_i\right] \\
&\quad+\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\frac{f_{T|X}(T_i+\delta|\boldsymbol{X}_i)}{f_{T|X}(T_i|\boldsymbol{X}_i)}\mu(T_i+\delta,\boldsymbol{X}_i)\right]
\end{aligned} \tag{83}
$$

$$
\begin{aligned}
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Bigg\{\pi_0(T_i-\delta,\boldsymbol{X}_i)\phi(T_i,\boldsymbol{X}_i,Y_i)-\pi_0(T_i,\boldsymbol{X}_i)\frac{f_{T|X}(T_i+\delta|\boldsymbol{X}_i)}{f_{T|X}(T_i|\boldsymbol{X}_i)}\mu(T_i+\delta,\boldsymbol{X}_i) \\
&+\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\frac{f_{T|X}(T_i+\delta|\boldsymbol{X}_i)}{f_{T|X}(T_i|\boldsymbol{X}_i)}\mu(T_i+\delta,\boldsymbol{X}_i)\middle|\boldsymbol{X}_i\right]-\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\frac{f_{T|X}(T_i+\delta|\boldsymbol{X}_i)}{f_{T|X}(T_i|\boldsymbol{X}_i)}\mu(T_i+\delta,\boldsymbol{X}_i)\right]
\end{aligned} \tag{84}
$$

$$+\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\frac{f_{T|X}(T_i+\delta|\boldsymbol{X}_i)}{f_{T|X}(T_i|\boldsymbol{X}_i)}\mu(T_i+\delta,\boldsymbol{X}_i)\middle|T_i\right]-\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\frac{f_{T|X}(T_i+\delta|\boldsymbol{X})}{f_{T|X}(T_i|\boldsymbol{X}_i)}\mu(T_i+\delta,\boldsymbol{X}_i)\right]\Bigg\}.$$

Using changing of variables, the first term of (83) can be written as follows:

$$\begin{aligned}&\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}\mu(t,\boldsymbol{x})\rho''\left(u_{K_1}^\top(t-\delta)\tilde{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)u_{K_1}^\top(t-\delta)\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})f_{X,T}(\boldsymbol{x},t)dxdt\\=&\sqrt{N}\int_{\mathcal{T}}\int_{\mathcal{X}}\left\{\frac{f_{T|X}(t+\delta|\boldsymbol{x})}{f_{T|X}(t|\boldsymbol{x})}\right\}\mu(t+\delta,\boldsymbol{x})\rho''\left(u_{K_1}^\top(t)\tilde{\Lambda}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right)u_{K_1}^\top(t)\hat{A}_{K_1\times K_2}v_{K_2}(\boldsymbol{x})dF_{X,T}(\boldsymbol{x},t).\end{aligned}$$

Using a similar argument of showing that (46)-(49) are all $o_p(1)$, we can show that (80)-(83) are all $o_p(1)$. By substituting $\phi(T,\boldsymbol{X},Y)=u_{K_1}(t_1)^\top\Phi_{K_1\times K_1}u_{K_1}(T)\pi_0(T,\boldsymbol{X})Y/\pi_0(T-\delta,\boldsymbol{X})^2$, we can obtain

$$\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot b_{1N}^{(1)}(t_1)=V_{t_1,t_0|t_0}^{-1/2}\cdot u_{K_1}^\top(t_1)\Phi_{K_1\times K_1}^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^N b_{1,K_1}(T_i,\boldsymbol{X}_i,Y_i)\right]+o_P(1),\qquad(85)$$

where

$$\begin{aligned}b_{1,K_1}(T_i,\boldsymbol{X}_i,Y_i)=&\frac{f_{T|X}(T_i+\delta|\boldsymbol{X})}{f_{T|X}(T_i|\boldsymbol{X}_i)}\frac{\pi_0(T_i,\boldsymbol{X}_i)^2}{\pi_0(T_i-\delta,\boldsymbol{X}_i)^2}\cdot\mathbb{E}[Y_i|T_i,\boldsymbol{X}_i]\cdot u_{K_1}(T_i)\\&-\mathbb{E}\left[\frac{f_{T|X}(T_i+\delta|\boldsymbol{X})}{f_{T|X}(T_i|\boldsymbol{X}_i)}\frac{\pi_0(T_i,\boldsymbol{X}_i)^2}{\pi_0(T_i-\delta,\boldsymbol{X}_i)^2}\cdot Y_i\cdot u_{K_1}(T_i)\middle|\boldsymbol{X}_i\right]\\&-\mathbb{E}\left[\frac{f_{T|X}(T_i+\delta|\boldsymbol{X})}{f_{T|X}(T_i|\boldsymbol{X}_i)}\frac{\pi_0(T_i,\boldsymbol{X}_i)^2}{\pi_0(T_i-\delta,\boldsymbol{X}_i)^2}\cdot Y_i\cdot u_{K_1}(T_i)\middle|T_i\right]\\&+\mathbb{E}\left[\frac{f_{T|X}(T_i+\delta|\boldsymbol{X})}{f_{T|X}(T_i|\boldsymbol{X}_i)}\frac{\pi_0(T_i,\boldsymbol{X}_i)^2}{\pi_0(T_i-\delta,\boldsymbol{X}_i)^2}\cdot Y_i\cdot u_{K_1}(T_i)\right].\end{aligned}$$

By combining (85), (76), and (77), we have

$$\begin{aligned}&\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot b_{1N}(t_1)\\=&\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot b_{1N}^{(1)}(t_1)+\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot b_{1N}^{(2)}(t_1)+\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot b_{1N}^{(3)}(t_1)\\=&V_{t_1,t_0|t_0}^{-1/2}\cdot u_{K_1}(t_1)^\top\Phi_{K_1\times K_1}^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^N b_{1,K_1}(T_i,\boldsymbol{X}_i,Y_i)\right]+o_P(1).\end{aligned}\qquad(86)$$

Similar to the proof of (40), we can show

$$\sqrt{N}V_{t_1,t_0|t_0}^{-1/2}\cdot b_{2N}(t_1)=u_{K_1}(t_1)^\top\Phi_{K_1\times K_1}^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^N b_{2,K_1}(T_i,\boldsymbol{X}_i,Y_i)\right]+o_P(1),\qquad(87)$$

where

$$b_{2,K_1}(T_i, \boldsymbol{X}_i, Y_i) = \frac{\pi_0(T_i, \boldsymbol{X}_i)}{\pi_0(T_i - \delta, \boldsymbol{X}_i)} \cdot Y_i \cdot u_{K_1}(T_i) - \mathbb{E}\left[\frac{\pi_0(T_i, \boldsymbol{X}_i)}{\pi_0(T_i - \delta, \boldsymbol{X}_i)} \cdot Y_i \cdot u_{K_1}(T_i) \middle| T_i, \boldsymbol{X}_i\right]$$
$$+ \mathbb{E}\left[\frac{\pi_0(T_i, \boldsymbol{X}_i)}{\pi_0(T_i - \delta, \boldsymbol{X}_i)} \cdot Y_i \cdot u_{K_1}(T_i) \middle| \boldsymbol{X}_i\right] - \mathbb{E}\left[\frac{\pi_0(T_i, \boldsymbol{X}_i)}{\pi_0(T_i - \delta, \boldsymbol{X}_i)} \cdot Y_i \cdot u_{K_1}(T_i)\right].$$

Therefore, by combining (75), (86) and (87), we can have that

$$\sqrt{N} V_{t_1,t_0|t_0}^{-1/2} \cdot \widehat{\theta}_{t_1|t_0} = V_{t_1,t_0|t_0}^{-1/2} \cdot u_{K_1}(t_1)^\top \Phi_{K_1 \times K_1}^{-1} \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{b_{1,K_1}(T_i, \boldsymbol{X}_i, Y_i) + b_{2,K_1}(T_i, \boldsymbol{X}_i, Y_i)\} + o_P(1). \tag{88}$$

By combining (74) and (88), we can obtain

$$\sqrt{N} V_{t_1,t_0|t_0}^{-1/2} \cdot \widehat{\theta}_{t_1,t_0|t_0} = \sqrt{N} V_{t_1,t_0|t_0}^{-1/2} \cdot \left\{\widehat{\theta}_{t_1|t_0} - \widehat{\theta}_{t_0|t_0}\right\}$$
$$= V_{t_1,t_0|t_0}^{-1/2} \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left\{u_{K_1}(t_1)^\top \Phi_{K_1 \times K_1}^{-1} \cdot \{b_{1,K_1}(T_i, \boldsymbol{X}_i, Y_i) + b_{2,K_1}(T_i, \boldsymbol{X}_i, Y_i)\}\right.$$
$$\left. - u_{K_1}(t_0)^\top \Phi_{K_1 \times K_1}^{-1} \cdot b_{3,K_1}(T_i, Y_i)\right\} + o_P(1),$$

which implies $\sqrt{N} V_{t_1,t_0|t_0}^{-1/2} \cdot \widehat{\theta}_{t_1,t_0|t_0} \xrightarrow{d} N(0,1)$ by Liapounov CLT.

6 Complete Monte Carlo Simulations

To evaluate the finite sample performance of the generalized optimization estimator, we conduct a simulation study on a continuous treatment. We present a simulation design in Section 6.1 and results in Section 6.2.

6.1 Simulation design

Let $\boldsymbol{X}_i = (X_{1i}, X_{2i})^\top$ be covariates, and assume that $\boldsymbol{X}_i \overset{i.i.d.}{\sim} N(0, I_2)$. Error terms are drawn mutually independently as $\xi_i \overset{i.i.d.}{\sim} N(0,1)$ and $\epsilon_i \overset{i.i.d.}{\sim} N(0,1)$. We consider four data generating processes (DGPs):

DGP-L1 $T = 1 + 0.2X_1 + \xi$ and $Y = 1 + X_1 + T + \epsilon$. ($X_2$ does not play any role, and X_1 affects T and Y linearly.)

DGP-NL1 $T = 0.1X_1^2 + \xi$ and $Y = X_1^2 + T + \epsilon$. (X_2 does not play any role, and X_1

affects T and Y non-linearly.)

DGP-L2 $T = 1 + 0.2\sum_{j=1}^{2} X_j + \xi$ and $Y = 1 + (1/2)\sum_{j=1}^{2} X_j + T + \epsilon$. ($X_1$ and X_2 affect T and Y linearly.)

DGP-NL2 $T = 0.1(\sum_{j=1}^{2} X_j)^2 + \xi$ and $Y = 1/2 + [(1/2)\sum_{j=1}^{2} X_j]^2 + T + \epsilon$. ($X_1$ and X_2 affect T and Y non-linearly.)

For each DGP, the true link function is $\mathbb{E}[Y(t)] = 1 + t$, a simple linear function with $\beta_1^* = \beta_2^* = 1$. Below we use a linear link function $g(T;\beta) = \beta_1 + \beta_2 T$, compute the generalized optimization estimator $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)^\top$, and examine its performance in terms of point and interval estimation.

To compute the generalized optimization estimator, two sieve basis functions $u_{K_1}(T)$ and $v_{K_2}(\boldsymbol{X})$ need to be specified. For $u_{K_1}(T)$, $K_1 \in \{2,3,4\} \equiv \mathbb{K}_1$ is considered:

$$u_2(T) = (1,T)^\top,\ u_3(T) = (1,T,T^2)^\top,\ u_4(T) = (1,T,T^2,T^3)^\top.$$

For $v_{K_2}(\boldsymbol{X})$, the choice set $\mathbb{K}_2$ depends on the number of covariates. For DGP-L1 and DGP-NL1, $K_2 \in \{2,3,4\} \equiv \mathbb{K}_2^1$ is considered:

$$v_2(X_1) = (1,X_1)^\top,\ v_3(X_1) = (1,X_1,X_1^2)^\top,\ v_4(X_1) = (1,X_1,X_1^2,X_1^3)^\top. \tag{89}$$

For DGP-L2 and DGP-NL2, $K_2 \in \{3,6,10\} \equiv \mathbb{K}_2^2$ is considered:

$$\begin{aligned} v_3(\boldsymbol{X}) &= (1,X_1,X_2)^\top, \\ v_6(\boldsymbol{X}) &= (1,X_1,X_2,X_1^2,X_2^2,X_1X_2)^\top, \\ v_{10}(\boldsymbol{X}) &= (1,X_1,X_2,X_1^2,X_2^2,X_1X_2,X_1^3,X_2^3,X_1^2X_2,X_1X_2^2)^\top. \end{aligned} \tag{90}$$

In addition to fixed pairs of $(K_1,K_2) \in \mathbb{K}_1 \times \mathbb{K}_2$, the data-driven selections described in the main paper Ai, Linton, Motegi, and Zhang (2020, Section 7) are employed. First, the (penalized) loss function approaches are implemented with the quadratic loss function $L\{Y - g(T;\boldsymbol{\beta})\} = (Y - \beta_1 - \beta_2 T)^2$. Second, the J-fold cross validation with $J \in \{5,10\}$ is implemented. For both approaches, the choice set is $\mathbb{K}_1 \times \mathbb{K}_2$.

We also compute Fong, Hazlett, and Imai's (2018) covariate balancing generalized propensity score estimator with a linear model specification and the quadratic loss function. The linear specification is correct under DGP-L1 and DGP-L2, while it is incorrect under DGP-NL1 and

DGP-NL2. Comparing our estimator and the parametric estimator of Fong, Hazlett, and Imai (2018) allows us to highlight the robustness of the former to non-linear DGPs. Fong, Hazlett, and Imai (2018) also propose a nonparametric estimator in their Section 3.3. In their simulation study, the parametric and nonparametric estimators exhibit similar performance for each DGP considered (Fong, Hazlett, and Imai, 2018, Figure 2). Hence, the present simulation study focuses on the parametric version of their estimator to save space.

Our proposed estimator and the parametric version of Fong, Hazlett, and Imai's (2018) estimator are computed in a simulated sample with size $N \in \{100, 500, 1000\}$, after which another sample is generated and both estimators are computed again. This exercise is repeated $M = 1000$ times.

To evaluate the performance of point estimation, the bias, standard deviation, and root mean squared error (RMSE) of $\hat{\beta}_1$ and $\hat{\beta}_2$ are calculated from (a subset of) $M = 1000$ simulations. In a small portion of the $M = 1000$ samples, $\bar{\pi}_N \equiv (1/N) \sum_{i=1}^{N} \hat{\pi}_K(T_i, \boldsymbol{X}_i)$, which should be equal to 1 in theory, takes a value far from 1 due to numerical instability in the computation of $\Lambda^*_{K_1 \times K_2}$. The numerical maximization with respect to Λ should lead to a global maximizer $\Lambda^*_{K_1 \times K_2}$ in theory, but optimizing the $K_1 \times K_2$ elements of Λ all at once is often hard in practice. Hence, we calculate the bias, standard deviation, and RMSE from Monte Carlo samples such that $\bar{\pi}_N \in [0.5, 2]$. There can be a few samples in which $\bar{\pi}_N \notin [0.5, 2]$, and these samples are simply discarded. (We admit that this computational problem becomes worse as the dimension of $\boldsymbol{X}$ becomes larger.)

To evaluate the finite sample performance of the interval estimation associated with the proposed method, we implement a bootstrap method with $B = 500$ iterations. In this method, we construct bootstrapped confidence intervals without using the asymptotic normality (see the main paper Ai, Linton, Motegi, and Zhang, 2020, Eq. (6.6)). For each of β_1 and β_2, we compute the 95% coverage probability and the average width of the bootstrapped confidence intervals across $M = 1000$ Monte Carlo samples. For simplicity, the dimensions of the sieve basis functions are fixed at $(K_1, K_2) = (2, 3)$ for DGP-L1 and DGP-NL1 and $(K_1, K_2) = (2, 6)$ for DGP-L2 and DGP-NL2 when the performance of the interval estimation is evaluated.

6.2 Simulation results

We discuss point estimation first, and then discuss variance estimation. See Tables 1-8 for the bias, standard deviation, and RMSE. In Figures 1-8, we draw bar charts that depict the share of (K_1, K_2) selected by each data-driven method. Under DGP-L1, the generalized optimization estimator (labeled as GOE) has sufficiently small RMSE for any fixed (K_1, K_2) (Tables 1-2).

It is not a surprising result since DGP-L1 has a simple linear structure. The data-driven methods often choose $(K_1^*, K_2^*) = (2, 2)$, the simplest possible approximation basis (Figures 1-2). The RMSE of the parametric version of the covariate balancing generalized propensity score estimator (labeled as CBGPS) is even smaller than the RMSE of GOE for β_1, and as small as the RMSE of GOE for β_2 (Tables 1-2). It is not surprising since CBGPS has a correct parametric specification under DGP-L1.

Under DGP-NL1, GOE dominates CBGPS. GOE leads to sufficiently small RMSE as long as $K_2 \geq 3$ (Tables 3-4). The relatively large RMSE for β_2 under $K_2 = 2$ suggests that X_1^2 needs to be included in $v_{K_2}(X_1)$ (see (89)). That is a reasonable result since DGP-NL1 has a quadratic structure. As expected, any data-driven method considered often selects pairs with $K_2 \geq 3$ (Figures 3-4). CBGPS, in contrast, fails with the bias in β_2 being around 0.2. The bias arises from the fact that the linear specification of CBGPS is incorrect under DGP-NL1. This result highlights that GOE performs well for both linear and nonlinear scenarios while CBGPS performs well for linear scenarios only.

The two-covariate scenarios yield similar implications to the single-covariate scenarios. Under DGP-L2, GOE with any fixed (K_1, K_2) has small RMSE (Tables 5-6). The data-driven methods often choose $(K_1^*, K_2^*) = (2, 3)$, the simplest possible approximation basis (Figures 5-6). The RMSE of CBGPS is even smaller than the RMSE of GOE for β_1 and as small as the RMSE of GOE for β_2 due to the linear structures of DGP-L2.

Under DGP-NL2, GOE with $K_2 \geq 6$ leads to small RMSE, and any data-driven method considered often selects pairs with $K_2 \geq 6$ (Tables 7-8 and Figures 7-8). CBGPS, in contrast, fails with substantial bias of around 0.18 in β_2. To summarize the point estimation, the generalized optimization estimator performs well in finite samples, and its performance is still good even when the true DGP is nonlinear; in contrast, the existing parametric estimator of Fong, Hazlett, and Imai (2018) is sensitive to model misspecification.

See Table 9 for results on interval estimation associated with GOE. For any parameter, DGP, and sample size, the 95% coverage probability is close enough to 0.95. The average width of the 95% confidence interval shrinks as the sample size grows. See β_1 under DGP-L1, for instance. The coverage probability is $\{0.957, 0.966, 0.941\}$ for $N \in \{100, 500, 1000\}$, respectively. Similarly, the average width of the confidence interval is $\{0.709, 0.311, 0.220\}$. These results indicate that the bootstrap method leads to sufficiently accurate interval estimation under both linear and non-linear DGPs.

Table 1: Simulation results on point estimation of intercept β_1 under DGP-L1 (truth: $\beta_1^* = 1$)

		$N = 100$			$N = 500$			$N = 1000$		
	(K_1, K_2)	Bias	Stdev	RMSE	Bias	Stdev	RMSE	Bias	Stdev	RMSE
GOE	$(2, 2)$	−0.002	0.183	0.183	0.000	0.082	0.082	0.002	0.056	0.056
GOE	$(2, 3)$	0.005	0.187	0.187	0.001	0.083	0.083	−0.001	0.055	0.055
GOE	$(2, 4)$	0.002	0.188	0.188	−0.000	0.083	0.083	0.001	0.055	0.055
GOE	$(3, 2)$	0.011	0.185	0.185	0.001	0.081	0.081	0.001	0.059	0.059
GOE	$(3, 3)$	−0.002	0.188	0.189	−0.008	0.085	0.085	−0.004	0.057	0.057
GOE	$(3, 4)$	−0.016	0.205	0.206	−0.011	0.085	0.085	−0.010	0.059	0.059
GOE	$(4, 2)$	0.006	0.192	0.192	−0.002	0.080	0.080	−0.000	0.058	0.058
GOE	$(4, 3)$	−0.017	0.203	0.204	−0.010	0.084	0.085	−0.009	0.060	0.061
GOE	$(4, 4)$	−0.022	0.224	0.225	−0.013	0.089	0.090	−0.014	0.061	0.063
GOE	MSE (none)	−0.005	0.202	0.203	−0.009	0.086	0.087	−0.005	0.059	0.059
GOE	MSE (add)	−0.012	0.194	0.195	−0.005	0.081	0.081	−0.008	0.059	0.059
GOE	MSE (multi)	−0.003	0.190	0.190	−0.001	0.081	0.081	−0.005	0.055	0.056
GOE	CV ($J = 5$)	0.011	0.178	0.178	0.003	0.081	0.081	0.002	0.055	0.055
GOE	CV ($J = 10$)	−0.005	0.190	0.190	0.006	0.080	0.080	0.005	0.057	0.057
CBGPS	-	−0.005	0.149	0.149	0.001	0.067	0.067	0.001	0.045	0.045

DGP-L1: $T = 1 + X_1 + \xi$ and $Y = 1 + X_1 + T + \epsilon$, where $X_1 \sim N(0, 1)$. "GOE" is the proposed generalized optimization estimator. K_1 and K_2 are the dimensions of the polynomials of T and X_1, respectively. "MSE (none)" signifies that we pick (K_1, K_2) that minimizes $\bar{L}(K_1, K_2) = N^{-1} \sum_{i=1}^{N} \hat{\pi}(T_i, X_{1i})[Y_i - g(T_i; \hat{\beta})]^2$. "MSE (add)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2(K_1 + K_2)/N) \times \bar{L}(K_1, K_2)$. "MSE (multi)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2K_1K_2/N) \times \bar{L}(K_1, K_2)$. "CV" signifies the J-fold cross validation with $J \in \{5, 10\}$. The choice set of (K_1, K_2) is the nine pairs listed in the table. "CBGPS" is Fong, Hazlett, and Imai's (2018) parametric covariate balancing generalized propensity score estimator. The sample size is $N \in \{100, 500, 1000\}$, and the number of Monte Carlo iterations is $M = 1000$.

Table 2: Simulation results on point estimation of slope β_2 under DGP-L1 (truth: $\beta_2^* = 1$)

		$N = 100$			$N = 500$			$N = 1000$		
	(K_1, K_2)	Bias	Stdev	RMSE	Bias	Stdev	RMSE	Bias	Stdev	RMSE
GOE	$(2,2)$	0.005	0.109	0.109	-0.000	0.047	0.047	-0.001	0.033	0.033
GOE	$(2,3)$	0.001	0.107	0.107	0.002	0.050	0.050	0.003	0.033	0.034
GOE	$(2,4)$	0.014	0.112	0.113	0.006	0.048	0.049	0.006	0.035	0.035
GOE	$(3,2)$	-0.000	0.108	0.108	-0.001	0.049	0.049	0.002	0.034	0.034
GOE	$(3,3)$	0.002	0.122	0.122	0.006	0.050	0.050	0.008	0.034	0.035
GOE	$(3,4)$	0.012	0.126	0.126	0.014	0.048	0.051	0.016	0.038	0.041
GOE	$(4,2)$	0.006	0.117	0.117	0.008	0.049	0.049	0.006	0.034	0.035
GOE	$(4,3)$	0.013	0.128	0.128	0.013	0.050	0.051	0.016	0.037	0.040
GOE	$(4,4)$	0.021	0.140	0.142	0.017	0.055	0.057	0.021	0.037	0.043
GOE	MSE (none)	0.003	0.131	0.131	0.008	0.052	0.053	0.008	0.035	0.036
GOE	MSE (add)	0.003	0.123	0.123	0.005	0.050	0.050	0.007	0.036	0.037
GOE	MSE (multi)	0.005	0.120	0.120	0.003	0.049	0.049	0.005	0.035	0.035
GOE	CV ($J = 5$)	-0.004	0.109	0.109	0.002	0.049	0.049	-0.000	0.033	0.033
GOE	CV ($J = 10$)	0.006	0.112	0.112	0.000	0.047	0.047	0.000	0.034	0.034
CBGPS	-	0.003	0.106	0.106	-0.001	0.049	0.049	-0.000	0.033	0.033

DGP-L1: $T = 1 + X_1 + \xi$ and $Y = 1 + X_1 + T + \epsilon$, where $X_1 \sim N(0, 1)$. "GOE" is the proposed generalized optimization estimator. K_1 and K_2 are the dimensions of the polynomials of T and X_1, respectively. "MSE (none)" signifies that we pick (K_1, K_2) that minimizes $\bar{L}(K_1, K_2) = N^{-1} \sum_{i=1}^{N} \hat{\pi}(T_i, X_{1i})[Y_i - g(T_i; \hat{\beta})]^2$. "MSE (add)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2(K_1 + K_2)/N) \times \bar{L}(K_1, K_2)$. "MSE (multi)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2K_1K_2/N) \times \bar{L}(K_1, K_2)$. "CV" signifies the J-fold cross validation with $J \in \{5, 10\}$. The choice set of (K_1, K_2) is the nine pairs listed in the table. "CBGPS" is Fong, Hazlett, and Imai's (2018) parametric covariate balancing generalized propensity score estimator. The sample size is $N \in \{100, 500, 1000\}$, and the number of Monte Carlo iterations is $M = 1000$.

Table 3: Simulation results on point estimation of intercept β_1 under DGP-NL1 (truth: $\beta_1^* = 1$)

		$N = 100$			$N = 500$			$N = 1000$		
	(K_1, K_2)	Bias	Stdev	RMSE	Bias	Stdev	RMSE	Bias	Stdev	RMSE
GOE	$(2, 2)$	-0.031	0.175	0.178	-0.024	0.078	0.082	-0.021	0.052	0.056
GOE	$(2, 3)$	0.002	0.176	0.176	0.004	0.079	0.079	-0.001	0.055	0.055
GOE	$(2, 4)$	-0.004	0.172	0.172	-0.002	0.080	0.080	0.000	0.056	0.056
GOE	$(3, 2)$	-0.040	0.167	0.172	-0.026	0.075	0.080	-0.022	0.053	0.057
GOE	$(3, 3)$	-0.019	0.180	0.181	0.002	0.081	0.081	0.003	0.055	0.055
GOE	$(3, 4)$	-0.025	0.185	0.187	-0.009	0.083	0.083	-0.002	0.057	0.057
GOE	$(4, 2)$	-0.054	0.185	0.192	-0.027	0.074	0.078	-0.025	0.054	0.059
GOE	$(4, 3)$	-0.024	0.192	0.194	0.002	0.084	0.084	0.000	0.056	0.056
GOE	$(4, 4)$	-0.044	0.190	0.195	-0.005	0.083	0.083	-0.001	0.059	0.059
GOE	MSE (none)	-0.056	0.177	0.186	-0.022	0.080	0.083	-0.016	0.055	0.057
GOE	MSE (add)	-0.066	0.187	0.198	-0.022	0.082	0.085	-0.018	0.056	0.059
GOE	MSE (multi)	-0.065	0.182	0.193	-0.023	0.082	0.085	-0.017	0.058	0.060
GOE	CV ($J = 5$)	-0.042	0.183	0.187	-0.016	0.080	0.082	-0.008	0.057	0.057
GOE	CV ($J = 10$)	-0.037	0.176	0.180	-0.012	0.077	0.078	-0.010	0.057	0.058
CBGPS	-	-0.035	0.179	0.182	-0.021	0.075	0.078	-0.021	0.053	0.057

DGP-NL1: $T = 0.1X_1^2 + \xi$ and $Y = X_1^2 + T + \epsilon$, where $X_1 \sim N(0, 1)$. "GOE" is the proposed generalized optimization estimator. K_1 and K_2 are the dimensions of the polynomials of T and X_1, respectively. "MSE (none)" signifies that we pick (K_1, K_2) that minimizes $\bar{L}(K_1, K_2) = N^{-1} \sum_{i=1}^{N} \hat{\pi}(T_i, X_{1i})[Y_i - g(T_i; \hat{\beta})]^2$. "MSE (add)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2(K_1 + K_2)/N) \times \bar{L}(K_1, K_2)$. "MSE (multi)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2K_1K_2/N) \times \bar{L}(K_1, K_2)$. "CV" signifies the J-fold cross validation with $J \in \{5, 10\}$. The choice set of (K_1, K_2) is the nine pairs listed in the table. "CBGPS" is Fong, Hazlett, and Imai's (2018) parametric covariate balancing generalized propensity score estimator. The sample size is $N \in \{100, 500, 1000\}$, and the number of Monte Carlo iterations is $M = 1000$.

Table 4: Simulation results on point estimation of slope β_2 under DGP-NL1 (truth: $\beta_2^* = 1$)

		$N = 100$			$N = 500$			$N = 1000$		
	(K_1, K_2)	Bias	Stdev	RMSE	Bias	Stdev	RMSE	Bias	Stdev	RMSE
GOE	$(2, 2)$	0.182	0.177	0.254	0.192	0.081	0.209	0.195	0.055	0.202
GOE	$(2, 3)$	0.004	0.104	0.104	−0.001	0.048	0.048	−0.000	0.033	0.033
GOE	$(2, 4)$	0.010	0.116	0.116	0.006	0.047	0.047	0.005	0.033	0.034
GOE	$(3, 2)$	0.179	0.189	0.261	0.190	0.081	0.207	0.194	0.057	0.202
GOE	$(3, 3)$	0.005	0.122	0.122	0.004	0.051	0.051	0.007	0.034	0.035
GOE	$(3, 4)$	0.014	0.133	0.133	0.011	0.051	0.052	0.011	0.037	0.039
GOE	$(4, 2)$	0.176	0.187	0.257	0.190	0.083	0.207	0.192	0.059	0.201
GOE	$(4, 3)$	0.020	0.133	0.134	0.014	0.051	0.053	0.012	0.038	0.040
GOE	$(4, 4)$	0.022	0.143	0.145	0.014	0.055	0.056	0.016	0.037	0.040
GOE	MSE (none)	0.017	0.141	0.142	0.021	0.069	0.072	0.020	0.057	0.060
GOE	MSE (add)	0.014	0.147	0.147	0.025	0.076	0.080	0.019	0.059	0.062
GOE	MSE (multi)	0.037	0.152	0.157	0.038	0.085	0.093	0.024	0.066	0.070
GOE	CV $(J = 5)$	0.107	0.170	0.201	0.081	0.106	0.133	0.069	0.097	0.119
GOE	CV $(J = 10)$	0.102	0.173	0.201	0.080	0.107	0.133	0.069	0.096	0.118
CBGPS	-	0.189	0.188	0.267	0.194	0.083	0.211	0.194	0.057	0.203

DGP-NL1: $T = 0.1X_1^2 + \xi$ and $Y = X_1^2 + T + \epsilon$, where $X_1 \sim N(0, 1)$. "GOE" is the proposed generalized optimization estimator. K_1 and K_2 are the dimensions of the polynomials of T and X_1, respectively. "MSE (none)" signifies that we pick (K_1, K_2) that minimizes $\bar{L}(K_1, K_2) = N^{-1} \sum_{i=1}^{N} \hat{\pi}(T_i, X_{1i})[Y_i - g(T_i; \hat{\beta})]^2$. "MSE (add)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2(K_1 + K_2)/N) \times \bar{L}(K_1, K_2)$. "MSE (multi)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2K_1K_2/N) \times \bar{L}(K_1, K_2)$. "CV" signifies the J-fold cross validation with $J \in \{5, 10\}$. The choice set of (K_1, K_2) is the nine pairs listed in the table. "CBGPS" is Fong, Hazlett, and Imai's (2018) parametric covariate balancing generalized propensity score estimator. The sample size is $N \in \{100, 500, 1000\}$, and the number of Monte Carlo iterations is $M = 1000$.

Table 5: Simulation results on point estimation of intercept β_1 under DGP-L2 (truth: $\beta_1^* = 1$)

		$N = 100$			$N = 500$			$N = 1000$		
	(K_1, K_2)	Bias	Stdev	RMSE	Bias	Stdev	RMSE	Bias	Stdev	RMSE
GOE	$(2, 3)$	−0.005	0.171	0.171	−0.000	0.073	0.073	−0.001	0.053	0.053
GOE	$(2, 6)$	−0.013	0.171	0.171	−0.010	0.081	0.082	−0.009	0.055	0.056
GOE	$(2, 10)$	−0.039	0.181	0.186	−0.034	0.078	0.085	−0.028	0.056	0.062
GOE	$(3, 3)$	−0.016	0.169	0.170	−0.002	0.075	0.075	0.000	0.057	0.057
GOE	$(3, 6)$	−0.027	0.195	0.197	−0.024	0.083	0.087	−0.026	0.061	0.066
GOE	$(3, 10)$	−0.032	0.202	0.205	−0.034	0.080	0.087	−0.030	0.058	0.065
GOE	$(4, 3)$	−0.014	0.179	0.180	−0.006	0.079	0.079	−0.008	0.056	0.056
GOE	$(4, 6)$	−0.036	0.207	0.210	−0.029	0.082	0.087	−0.030	0.059	0.066
GOE	$(4, 10)$	−0.032	0.211	0.213	−0.030	0.082	0.088	−0.025	0.059	0.064
GOE	MSE (none)	−0.038	0.210	0.213	−0.015	0.083	0.085	−0.015	0.058	0.060
GOE	MSE (add)	−0.011	0.191	0.192	−0.010	0.080	0.081	−0.012	0.057	0.058
GOE	MSE (multi)	−0.008	0.184	0.184	−0.000	0.076	0.076	−0.003	0.056	0.056
GOE	CV ($J = 5$)	0.001	0.174	0.174	−0.003	0.074	0.074	−0.004	0.054	0.054
GOE	CV ($J = 10$)	−0.003	0.169	0.169	0.001	0.078	0.078	−0.005	0.053	0.053
CBGPS	-	0.000	0.157	0.157	−0.001	0.067	0.067	0.002	0.049	0.049

DGP-L2: $T = 1 + 0.2\sum_{j=1}^{2} X_j + \xi$ and $Y = 1 + (1/2)\sum_{j=1}^{2} X_j + T + \epsilon$, where $X_1, X_2 \overset{i.i.d.}{\sim} N(0,1)$. "GOE" is the proposed generalized optimization estimator. K_1 and K_2 are the dimensions of the polynomials of T and $\boldsymbol{X} = (X_1, X_2)^\top$, respectively. "MSE (none)" signifies that we pick (K_1, K_2) that minimizes $\bar{L}(K_1, K_2) = N^{-1}\sum_{i=1}^{N} \hat{\pi}(T_i, \boldsymbol{X})[Y_i - g(T_i; \hat{\beta})]^2$. "MSE (add)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2(K_1 + K_2)/N) \times \bar{L}(K_1, K_2)$. "MSE (multi)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2K_1K_2/N) \times \bar{L}(K_1, K_2)$. "CV" signifies the J-fold cross validation with $J \in \{5, 10\}$. The choice set of (K_1, K_2) is the nine pairs listed in the table. "CBGPS" is Fong, Hazlett, and Imai's (2018) parametric covariate balancing generalized propensity score estimator. The sample size is $N \in \{100, 500, 1000\}$, and the number of Monte Carlo iterations is $M = 1000$.

Table 6: Simulation results on point estimation of slope β_2 under DGP-L2 (truth: $\beta_2^* = 1$)

		$N = 100$			$N = 500$			$N = 1000$		
	(K_1, K_2)	Bias	Stdev	RMSE	Bias	Stdev	RMSE	Bias	Stdev	RMSE
GOE	$(2, 3)$	0.001	0.108	0.108	0.002	0.047	0.047	0.000	0.035	0.035
GOE	$(2, 6)$	0.033	0.118	0.122	0.023	0.052	0.057	0.025	0.037	0.045
GOE	$(2, 10)$	0.051	0.128	0.138	0.042	0.052	0.067	0.040	0.038	0.055
GOE	$(3, 3)$	0.019	0.116	0.117	0.019	0.052	0.056	0.015	0.038	0.041
GOE	$(3, 6)$	0.023	0.133	0.135	0.029	0.054	0.062	0.031	0.040	0.050
GOE	$(3, 10)$	0.037	0.135	0.140	0.040	0.052	0.065	0.037	0.038	0.053
GOE	$(4, 3)$	0.024	0.122	0.125	0.021	0.052	0.056	0.023	0.037	0.044
GOE	$(4, 6)$	0.028	0.141	0.144	0.035	0.057	0.066	0.039	0.038	0.054
GOE	$(4, 10)$	0.030	0.141	0.144	0.029	0.055	0.062	0.029	0.040	0.049
GOE	MSE (none)	0.030	0.146	0.149	0.021	0.057	0.060	0.022	0.040	0.046
GOE	MSE (add)	0.010	0.127	0.127	0.014	0.054	0.056	0.018	0.040	0.044
GOE	MSE (multi)	0.014	0.120	0.121	0.007	0.050	0.050	0.010	0.038	0.039
GOE	CV ($J = 5$)	0.003	0.113	0.113	0.005	0.052	0.052	0.006	0.037	0.038
GOE	CV ($J = 10$)	0.012	0.117	0.118	0.004	0.051	0.051	0.007	0.037	0.038
CBGPS	-	0.001	0.111	0.111	0.001	0.049	0.049	−0.001	0.035	0.035

DGP-L2: $T = 1+0.2\sum_{j=1}^{2} X_j+\xi$ and $Y = 1+(1/2)\sum_{j=1}^{2} X_j+T+\epsilon$, where $X_1, X_2 \overset{i.i.d.}{\sim} N(0,1)$. "GOE" is the proposed generalized optimization estimator. K_1 and K_2 are the dimensions of the polynomials of T and $\boldsymbol{X} = (X_1, X_2)^\top$, respectively. "MSE (none)" signifies that we pick (K_1, K_2) that minimizes $\bar{L}(K_1, K_2) = N^{-1}\sum_{i=1}^{N} \hat{\pi}(T_i, \boldsymbol{X})[Y_i - g(T_i; \hat{\beta})]^2$. "MSE (add)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2(K_1 + K_2)/N) \times \bar{L}(K_1, K_2)$. "MSE (multi)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2K_1K_2/N) \times \bar{L}(K_1, K_2)$. "CV" signifies the J-fold cross validation with $J \in \{5, 10\}$. The choice set of (K_1, K_2) is the nine pairs listed in the table. "CBGPS" is Fong, Hazlett, and Imai's (2018) parametric covariate balancing generalized propensity score estimator. The sample size is $N \in \{100, 500, 1000\}$, and the number of Monte Carlo iterations is $M = 1000$.

Table 7: Simulation results on point estimation of intercept β_1 under DGP-NL2 (truth: $\beta_1^* = 1$)

		$N = 100$			$N = 500$			$N = 1000$		
	(K_1, K_2)	Bias	Stdev	RMSE	Bias	Stdev	RMSE	Bias	Stdev	RMSE
GOE	$(2, 3)$	−0.043	0.118	0.126	−0.037	0.053	0.065	−0.038	0.036	0.053
GOE	$(2, 6)$	−0.010	0.135	0.135	0.004	0.060	0.060	0.004	0.043	0.043
GOE	$(2, 10)$	−0.030	0.132	0.136	−0.008	0.057	0.058	−0.007	0.043	0.044
GOE	$(3, 3)$	−0.052	0.125	0.136	−0.041	0.053	0.067	−0.037	0.039	0.053
GOE	$(3, 6)$	−0.030	0.137	0.140	−0.007	0.060	0.060	−0.005	0.045	0.045
GOE	$(3, 10)$	−0.039	0.141	0.147	−0.017	0.060	0.062	−0.011	0.043	0.045
GOE	$(4, 3)$	−0.052	0.128	0.139	−0.038	0.055	0.067	−0.039	0.037	0.054
GOE	$(4, 6)$	−0.035	0.142	0.146	−0.012	0.061	0.062	−0.011	0.043	0.045
GOE	$(4, 10)$	−0.048	0.162	0.169	−0.018	0.061	0.063	−0.015	0.043	0.045
GOE	MSE (none)	−0.061	0.149	0.161	−0.025	0.057	0.063	−0.024	0.041	0.048
GOE	MSE (add)	−0.062	0.140	0.153	−0.031	0.058	0.066	−0.025	0.043	0.049
GOE	MSE (multi)	−0.046	0.133	0.140	−0.028	0.057	0.063	−0.022	0.042	0.048
GOE	CV $(J = 5)$	−0.041	0.130	0.136	−0.027	0.060	0.065	−0.019	0.045	0.049
GOE	CV $(J = 10)$	−0.050	0.121	0.131	−0.027	0.059	0.065	−0.021	0.041	0.049
CBGPS	-	−0.051	0.127	0.137	−0.038	0.053	0.065	−0.039	0.038	0.055

DGP-NL2: $T = 0.1(\sum_{j=1}^{2} X_j)^2 + \xi$ and $Y = 1/2 + [(1/2)\sum_{j=1}^{2} X_j]^2 + T + \epsilon$, where $X_1, X_2 \overset{i.i.d.}{\sim} N(0, 1)$. "GOE" is the proposed generalized optimization estimator. K_1 and K_2 are the dimensions of the polynomials of T and $\boldsymbol{X} = (X_1, X_2)^\top$, respectively. "MSE (none)" signifies that we pick (K_1, K_2) that minimizes $\bar{L}(K_1, K_2) = N^{-1} \sum_{i=1}^{N} \hat{\pi}(T_i, \boldsymbol{X})[Y_i - g(T_i; \hat{\beta})]^2$. "MSE (add)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2(K_1 + K_2)/N) \times \bar{L}(K_1, K_2)$. "MSE (multi)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2K_1K_2/N) \times \bar{L}(K_1, K_2)$. "CV" signifies the J-fold cross validation with $J \in \{5, 10\}$. The choice set of (K_1, K_2) is the nine pairs listed in the table. "CBGPS" is Fong, Hazlett, and Imai's (2018) parametric covariate balancing generalized propensity score estimator. The sample size is $N \in \{100, 500, 1000\}$, and the number of Monte Carlo iterations is $M = 1000$.

Table 8: Simulation results on point estimation of slope β_2 under DGP-NL2 (truth: $\beta_2^* = 1$)

		$N = 100$			$N = 500$			$N = 1000$		
	(K_1, K_2)	Bias	Stdev	RMSE	Bias	Stdev	RMSE	Bias	Stdev	RMSE
GOE	$(2, 3)$	0.172	0.129	0.215	0.185	0.060	0.194	0.184	0.040	0.188
GOE	$(2, 6)$	0.031	0.117	0.121	0.026	0.054	0.059	0.026	0.037	0.045
GOE	$(2, 10)$	0.047	0.131	0.139	0.041	0.054	0.067	0.039	0.037	0.053
GOE	$(3, 3)$	0.163	0.131	0.209	0.180	0.060	0.189	0.181	0.041	0.186
GOE	$(3, 6)$	0.027	0.132	0.134	0.031	0.054	0.062	0.031	0.039	0.050
GOE	$(3, 10)$	0.044	0.139	0.146	0.036	0.053	0.063	0.036	0.040	0.054
GOE	$(4, 3)$	0.162	0.133	0.210	0.175	0.059	0.185	0.180	0.040	0.184
GOE	$(4, 6)$	0.038	0.133	0.138	0.034	0.055	0.064	0.033	0.040	0.052
GOE	$(4, 10)$	0.029	0.149	0.151	0.028	0.054	0.061	0.030	0.037	0.047
GOE	MSE (none)	0.036	0.140	0.145	0.032	0.058	0.066	0.032	0.044	0.054
GOE	MSE (add)	0.055	0.152	0.162	0.043	0.071	0.083	0.039	0.052	0.065
GOE	MSE (multi)	0.106	0.140	0.175	0.076	0.083	0.113	0.054	0.068	0.087
GOE	CV $(J = 5)$	0.126	0.138	0.187	0.100	0.089	0.134	0.083	0.079	0.115
GOE	CV $(J = 10)$	0.131	0.137	0.190	0.102	0.090	0.136	0.082	0.079	0.114
CBGPS	-	0.176	0.136	0.223	0.184	0.058	0.193	0.184	0.043	0.189

DGP-NL2: $T = 0.1(\sum_{j=1}^{2} X_j)^2 + \xi$ and $Y = 1/2 + [(1/2)\sum_{j=1}^{2} X_j]^2 + T + \epsilon$, where $X_1, X_2 \overset{i.i.d.}{\sim} N(0, 1)$. "GOE" is the proposed generalized optimization estimator. K_1 and K_2 are the dimensions of the polynomials of T and $\boldsymbol{X} = (X_1, X_2)^\top$, respectively. "MSE (none)" signifies that we pick (K_1, K_2) that minimizes $\bar{L}(K_1, K_2) = N^{-1}\sum_{i=1}^{N} \hat{\pi}(T_i, \boldsymbol{X})[Y_i - g(T_i; \hat{\beta})]^2$. "MSE (add)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2(K_1 + K_2)/N) \times \bar{L}(K_1, K_2)$. "MSE (multi)" signifies that we pick (K_1, K_2) that minimizes $(1 + 2K_1K_2/N) \times \bar{L}(K_1, K_2)$. "CV" signifies the J-fold cross validation with $J \in \{5, 10\}$. The choice set of (K_1, K_2) is the nine pairs listed in the table. "CBGPS" is Fong, Hazlett, and Imai's (2018) parametric covariate balancing generalized propensity score estimator. The sample size is $N \in \{100, 500, 1000\}$, and the number of Monte Carlo iterations is $M = 1000$.

Table 9: Simulation results on interval estimation (generalized optimization estimator)

Intercept β_1 (true value: $\beta_1^* = 1$)

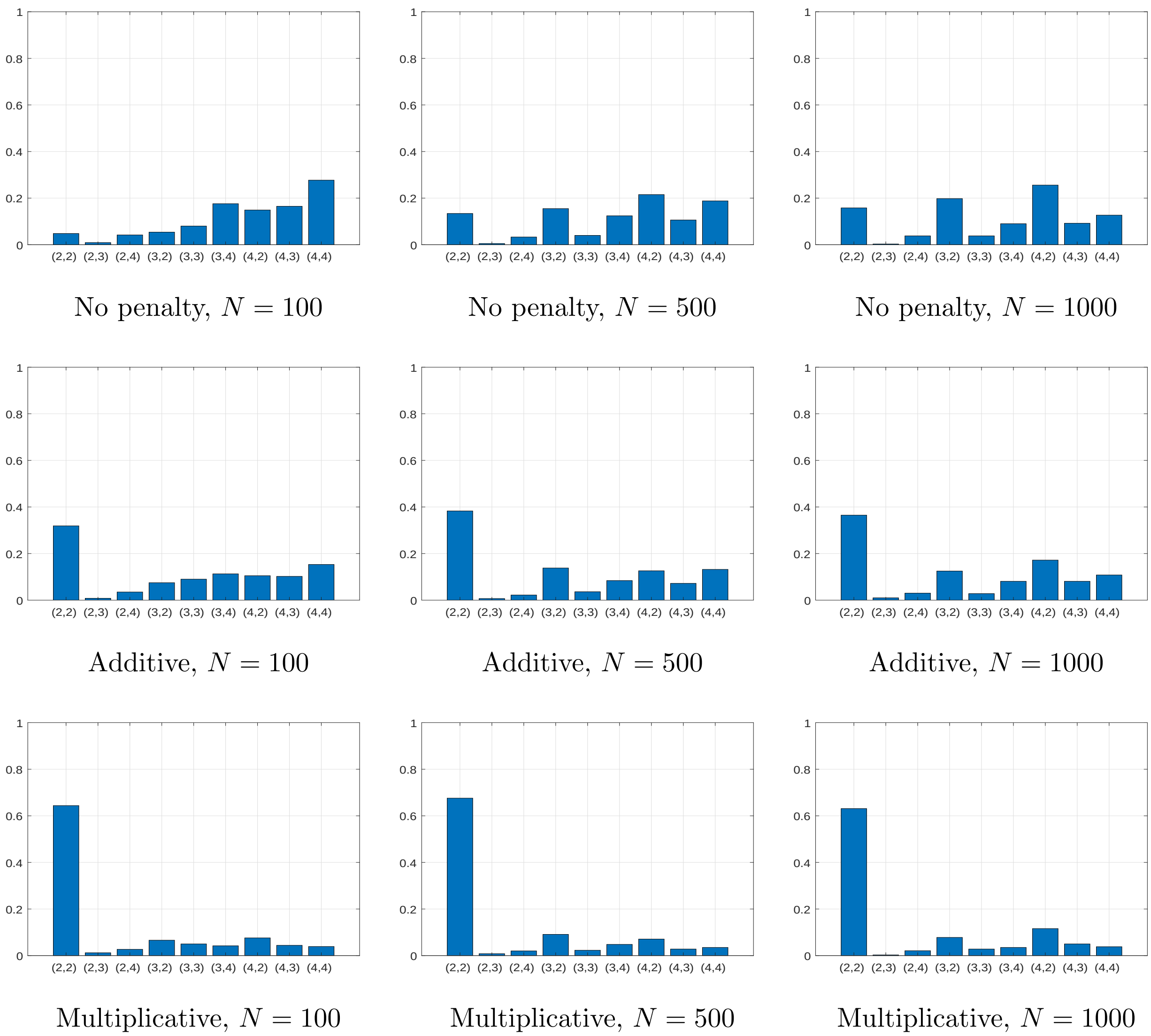
DGP	DGP-L1		DGP-NL1		DGP-L2		DGP-NL2	
(K_1, K_2)	$(2,3)$		$(2,3)$		$(2,6)$		$(2,6)$	
	CP95	AveW	CP95	AveW	CP95	AveW	CP95	AveW
$N = 100$	0.957	0.709	0.944	0.677	0.967	0.722	0.969	0.537
$N = 500$	0.966	0.311	0.942	0.305	0.960	0.304	0.953	0.236
$N = 1000$	0.941	0.220	0.955	0.217	0.959	0.221	0.967	0.171

Slope β_2 (true value: $\beta_2^* = 1$)

DGP	DGP-L1		DGP-NL1		DGP-L2		DGP-NL2	
(K_1, K_2)	$(2,3)$		$(2,3)$		$(2,6)$		$(2,6)$	
	CP95	AveW	CP95	AveW	CP95	AveW	CP95	AveW
$N = 100$	0.940	0.428	0.956	0.422	0.956	0.494	0.965	0.498
$N = 500$	0.947	0.184	0.950	0.180	0.945	0.205	0.937	0.208
$N = 1000$	0.933	0.130	0.957	0.128	0.929	0.149	0.934	0.153

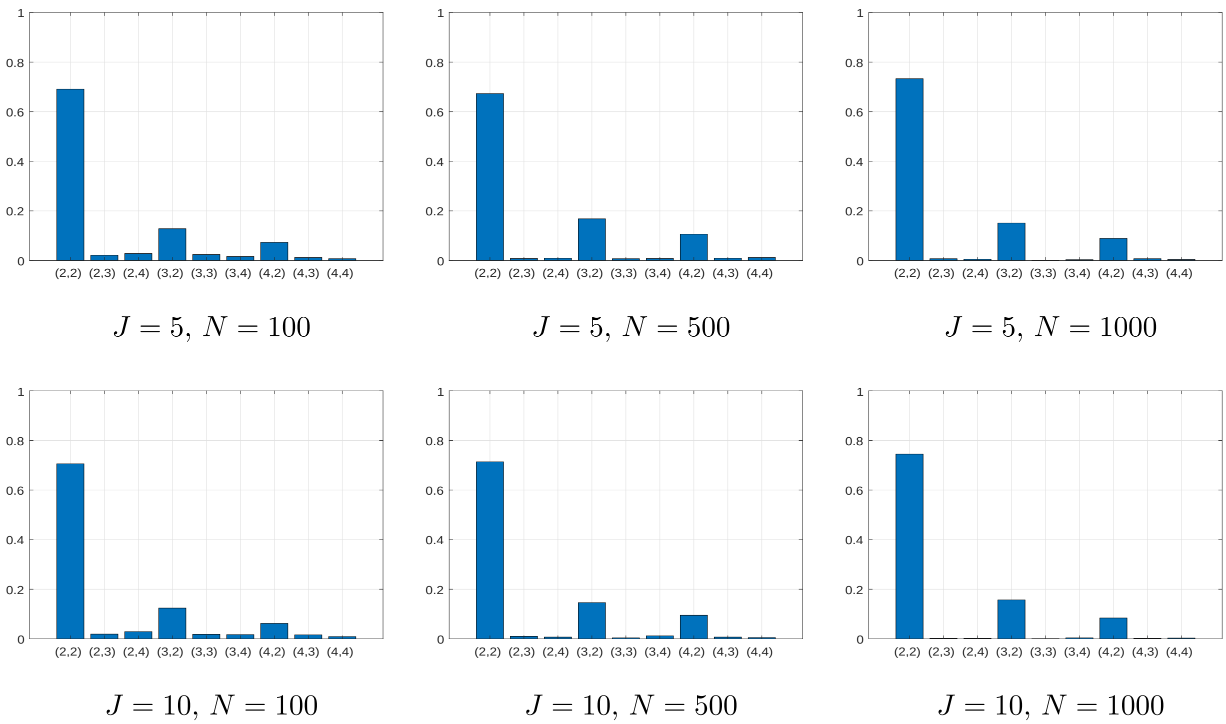
In this table, simulation results on the interval estimation associated with the generalized optimization estimator are presented. 95% confidence intervals on the target parameters (β_1, β_2) are constructed via the bootstrap with $B = 500$ iterations. K_1 and K_2 are the dimensions of the polynomials of T and covariate(s) $\boldsymbol{X}$, respectively. Under DGP-L1 and DGP-NL1, there is only one covariate X_1 and $(K_1, K_2) = (2,3)$. Under DGP-L2 and DGP-NL2, there are two covariates $\boldsymbol{X} = (X_1, X_2)^\top$ and $(K_1, K_2) = (2,6)$. Under L1 and L2, T and Y depend linearly on $\boldsymbol{X}$. Under NL1 and NL2, T and Y depend non-linearly on $\boldsymbol{X}$. "CP95" signifies the 95% coverage probability, while "AveW" signifies the average width of the confidence intervals across $M = 1000$ Monte Carlo samples.

Figure 1: Share of (K_1, K_2) selected under DGP-L1 (MSE criteria)



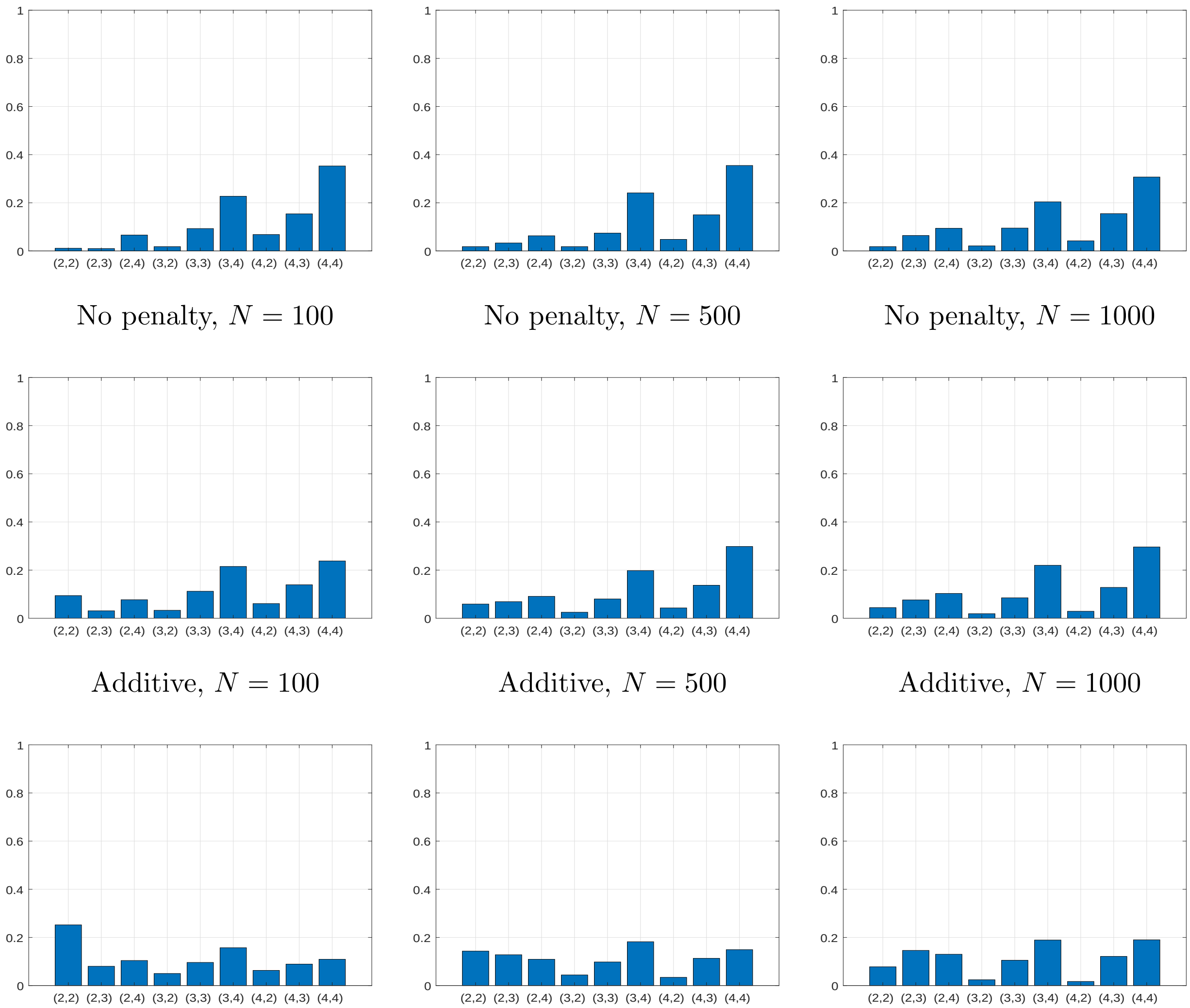
DGP-L1: $T = 1 + X_1 + \xi$ and $Y = 1 + X_1 + T + \epsilon$, where $X_1 \sim N(0, 1)$. K_1 and K_2 are the dimensions of the polynomials of T and X_1, respectively. "No penalty" signifies that we pick (K_1, K_2) that minimizes $\bar{L}(K_1, K_2) = N^{-1} \sum_{i=1}^{N} \hat{\pi}(T_i, X_{1i})[Y_i - g(T_i; \hat{\beta})]^2$. "Additive" signifies that we pick (K_1, K_2) that minimizes $(1 + 2(K_1 + K_2)/N) \times \bar{L}(K_1, K_2)$. "Multiplicative" signifies that we pick (K_1, K_2) that minimizes $(1 + 2K_1K_2/N) \times \bar{L}(K_1, K_2)$. In this figure, we plot the empirical probability of selecting each pair across $M = 1000$ Monte Carlo samples. The choice set of (K_1, K_2) is the nine pairs put on the horizontal axis.

Figure 2: Share of (K_1, K_2) selected under DGP-L1 (J-fold cross validation)



DGP-L1: $T = 1 + X_1 + \xi$ and $Y = 1 + X_1 + T + \epsilon$, where $X_1 \sim N(0, 1)$. K_1 and K_2 are the dimensions of the polynomials of T and X_1, respectively. We pick (K_1, K_2) that minimizes the loss function of the J-fold cross validation with $J \in \{5, 10\}$. In this figure, we plot the empirical probability of selecting each pair across $M = 1000$ Monte Carlo samples. The choice set of (K_1, K_2) is the nine pairs put on the horizontal axis.

Figure 3: Share of (K_1, K_2) selected under DGP-NL1 (MSE criteria)



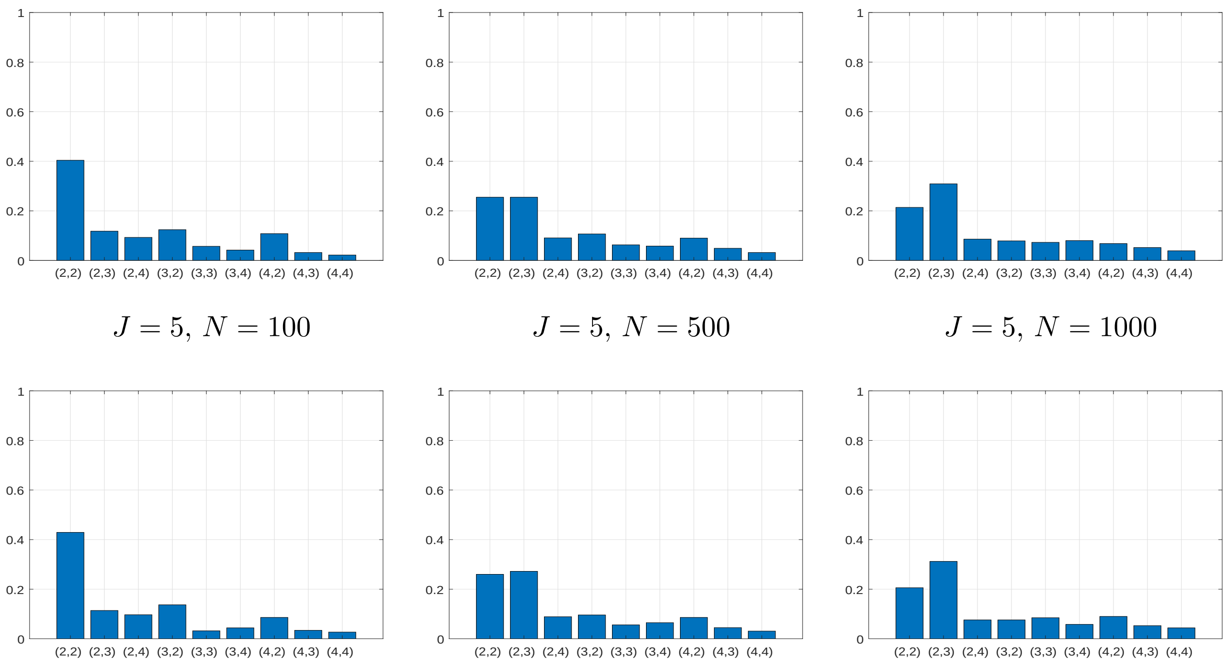
No penalty, $N = 100$ — No penalty, $N = 500$ — No penalty, $N = 1000$

Additive, $N = 100$ — Additive, $N = 500$ — Additive, $N = 1000$

Multiplicative, $N = 100$ — Multiplicative, $N = 500$ — Multiplicative, $N = 1000$

DGP-NL1: $T = 0.1X_1^2 + \xi$ and $Y = X_1^2 + T + \epsilon$, where $X_1 \sim N(0,1)$. K_1 and K_2 are the dimensions of the polynomials of T and X_1, respectively. “No penalty” signifies that we pick (K_1, K_2) that minimizes $\bar{L}(K_1, K_2) = N^{-1}\sum_{i=1}^{N} \hat{\pi}(T_i, X_{1i})[Y_i - g(T_i; \hat{\beta})]^2$. “Additive” signifies that we pick (K_1, K_2) that minimizes $(1 + 2(K_1 + K_2)/N) \times \bar{L}(K_1, K_2)$. “Multiplicative” signifies that we pick (K_1, K_2) that minimizes $(1 + 2K_1K_2/N) \times \bar{L}(K_1, K_2)$. In this figure, we plot the empirical probability of selecting each pair across $M = 1000$ Monte Carlo samples. The choice set of (K_1, K_2) is the nine pairs put on the horizontal axis.

Figure 4: Share of (K_1, K_2) selected under DGP-NL1 (J-fold cross validation)

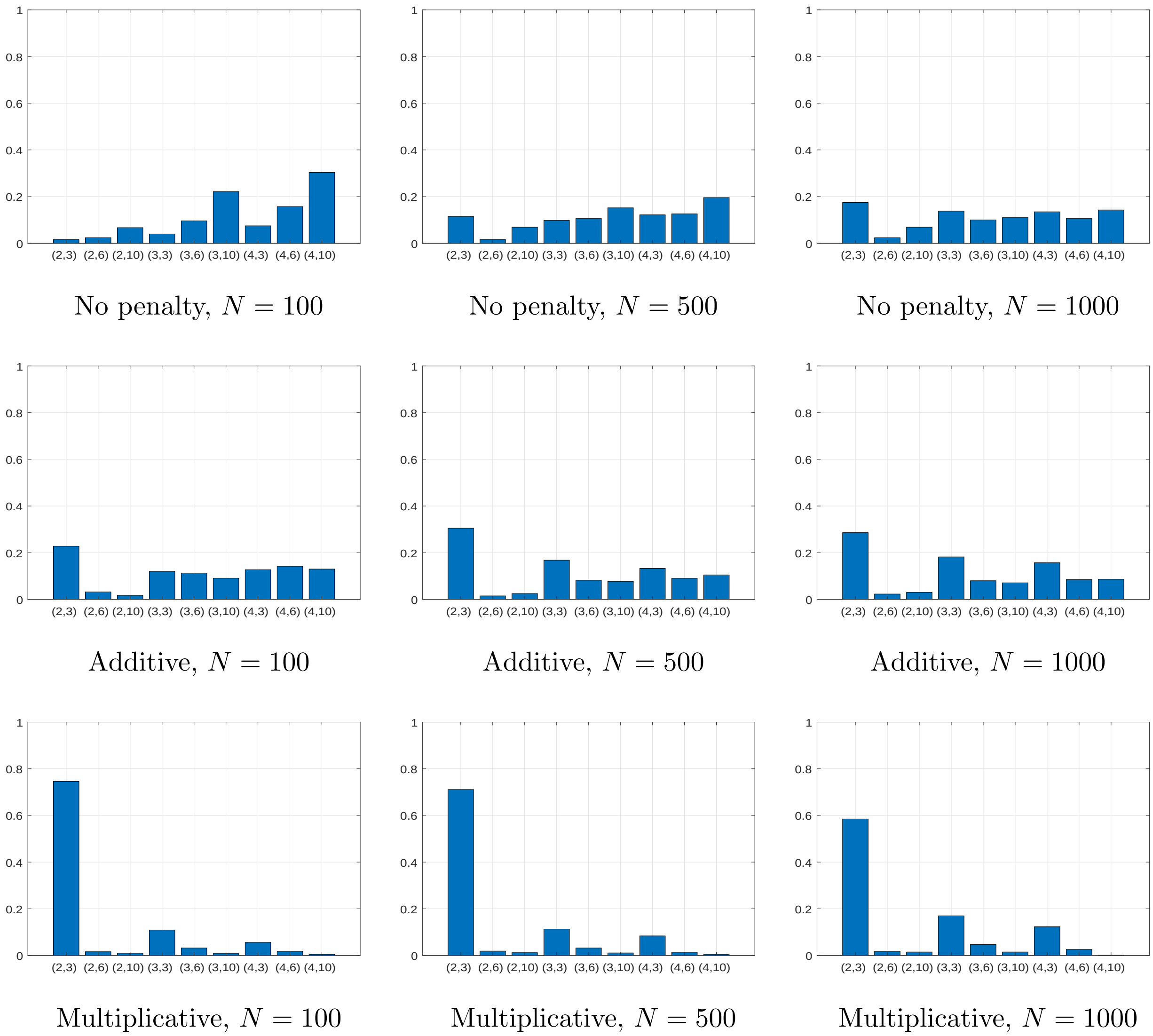


$J = 5,\ N = 100$ $\quad$ $J = 5,\ N = 500$ $\quad$ $J = 5,\ N = 1000$

$J = 10,\ N = 100$ $\quad$ $J = 10,\ N = 500$ $\quad$ $J = 10,\ N = 1000$

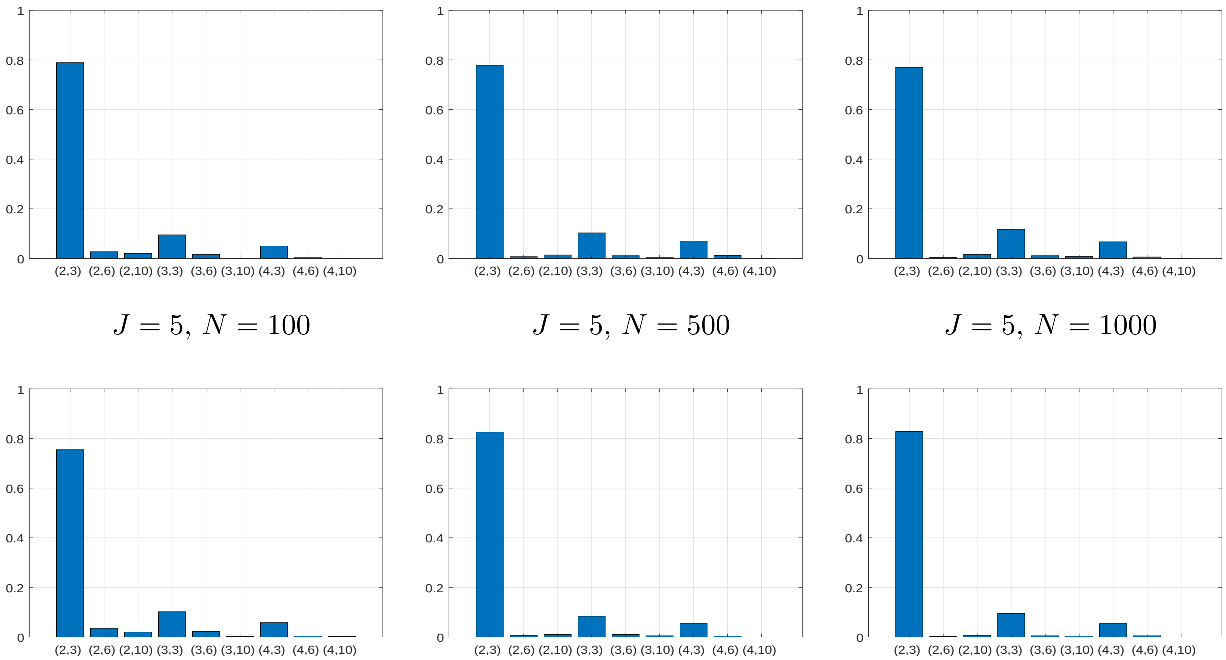
DGP-NL1: $T = 0.1X_1^2 + \xi$ and $Y = X_1^2 + T + \epsilon$, where $X_1 \sim N(0, 1)$. K_1 and K_2 are the dimensions of the polynomials of T and X_1, respectively. We pick (K_1, K_2) that minimizes the loss function of the J-fold cross validation with $J \in \{5, 10\}$. In this figure, we plot the empirical probability of selecting each pair across $M = 1000$ Monte Carlo samples. The choice set of (K_1, K_2) is the nine pairs put on the horizontal axis.

Figure 5: Share of (K_1, K_2) selected under DGP-L2 (MSE criteria)



DGP-L2: $T = 1 + 0.2\sum_{j=1}^{2} X_j + \xi$ and $Y = 1 + (1/2)\sum_{j=1}^{2} X_j + T + \epsilon$, where $X_1, X_2 \overset{i.i.d.}{\sim} N(0,1)$. K_1 and K_2 are the dimensions of the polynomials of T and $\boldsymbol{X} = (X_1, X_2)^\top$, respectively. "No penalty" signifies that we pick (K_1, K_2) that minimizes $\bar{L}(K_1, K_2) = N^{-1}\sum_{i=1}^{N} \hat{\pi}(T_i, \boldsymbol{X}_i)[Y_i - g(T_i; \hat{\beta})]^2$. "Additive" signifies that we pick (K_1, K_2) that minimizes $(1 + 2(K_1 + K_2)/N) \times \bar{L}(K_1, K_2)$. "Multiplicative" signifies that we pick (K_1, K_2) that minimizes $(1 + 2K_1K_2/N) \times \bar{L}(K_1, K_2)$. In this figure, we plot the empirical probability of selecting each pair across $M = 1000$ Monte Carlo samples. The choice set of (K_1, K_2) is the nine pairs put on the horizontal axis.

Figure 6: Share of (K_1, K_2) selected under DGP-L2 (J-fold cross validation)

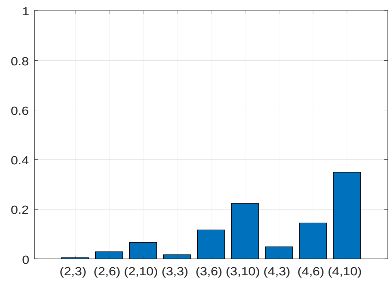


$J = 5,\ N = 100$ | $J = 5,\ N = 500$ | $J = 5,\ N = 1000$

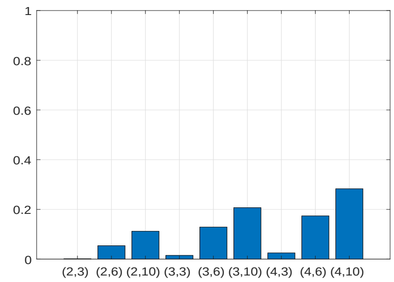
$J = 10,\ N = 100$ | $J = 10,\ N = 500$ | $J = 10,\ N = 1000$

DGP-L2: $T = 1 + 0.2 \sum_{j=1}^{2} X_j + \xi$ and $Y = 1 + (1/2) \sum_{j=1}^{2} X_j + T + \epsilon$, where $X_1, X_2 \overset{i.i.d.}{\sim} N(0, 1)$. K_1 and K_2 are the dimensions of the polynomials of T and $\boldsymbol{X} = (X_1, X_2)^\top$, respectively. We pick (K_1, K_2) that minimizes the loss function of the J-fold cross validation with $J \in \{5, 10\}$. In this figure, we plot the empirical probability of selecting each pair across $M = 1000$ Monte Carlo samples. The choice set of (K_1, K_2) is the nine pairs put on the horizontal axis.

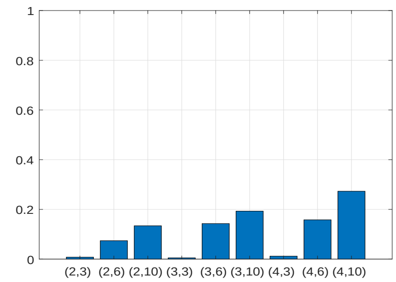
Figure 7: Share of (K_1, K_2) selected under DGP-NL2 (MSE criteria)



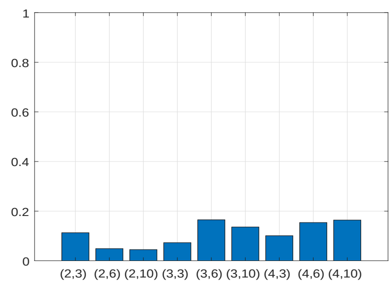
No penalty, $N = 100$



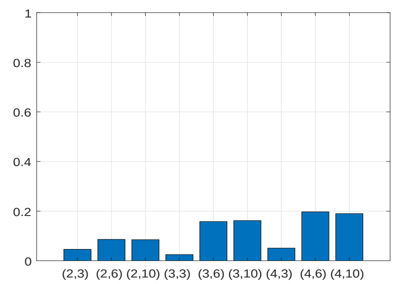
No penalty, $N = 500$



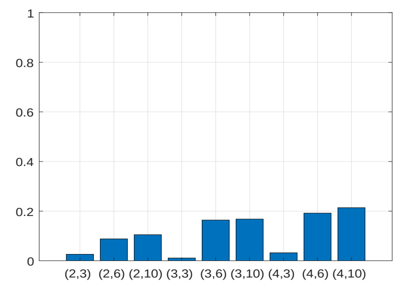
No penalty, $N = 1000$



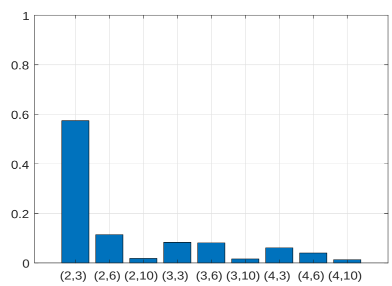
Additive, $N = 100$



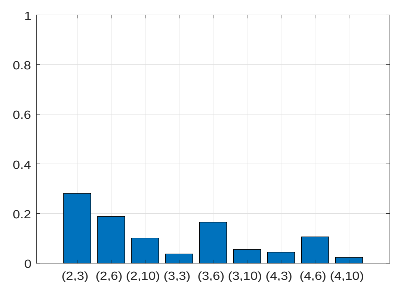
Additive, $N = 500$



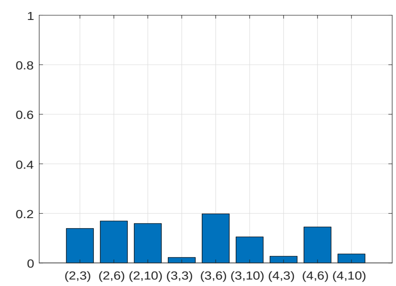
Additive, $N = 1000$



Multiplicative, $N = 100$



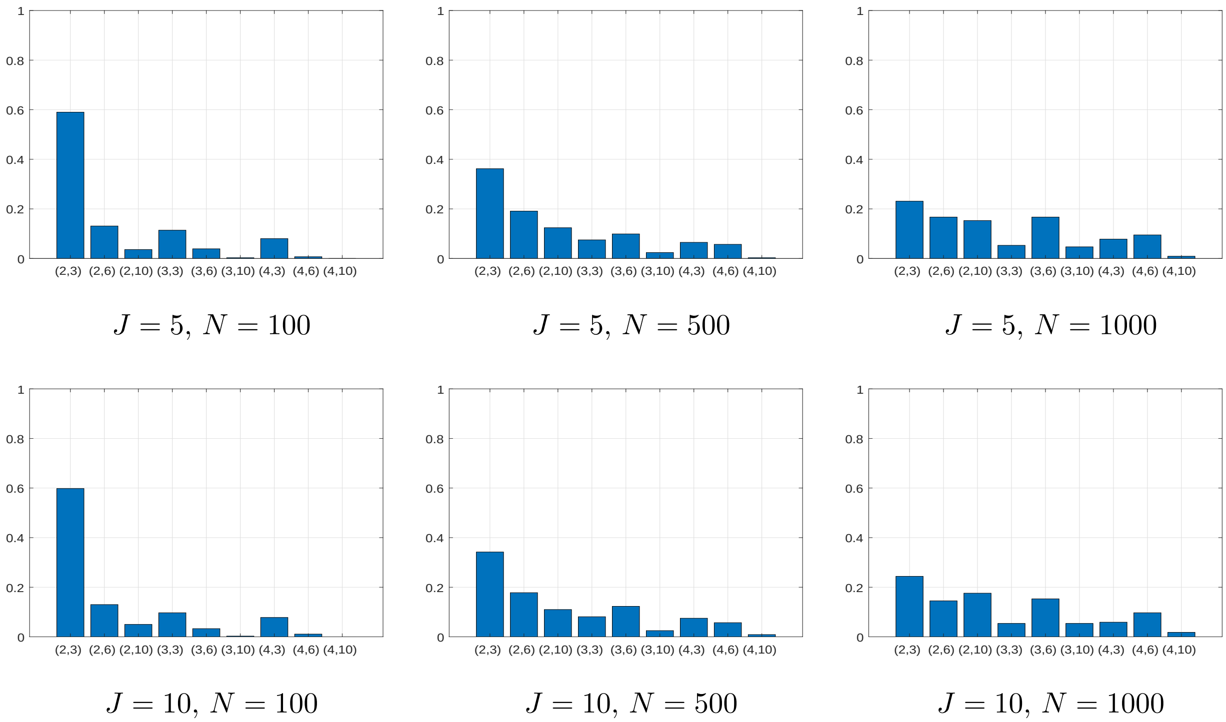
Multiplicative, $N = 500$



Multiplicative, $N = 1000$

DGP-NL2: $T = 0.1(\sum_{j=1}^{2} X_j)^2 + \xi$ and $Y = 1/2 + [(1/2)\sum_{j=1}^{2} X_j]^2 + T + \epsilon$, where $X_1, X_2 \overset{i.i.d.}{\sim} N(0,1)$. K_1 and K_2 are the dimensions of the polynomials of T and $\boldsymbol{X} = (X_1, X_2)^\top$, respectively. "No penalty" signifies that we pick (K_1, K_2) that minimizes $\bar{L}(K_1, K_2) = N^{-1}\sum_{i=1}^{N} \hat{\pi}(T_i, \boldsymbol{X}_i)[Y_i - g(T_i; \hat{\beta})]^2$. "Additive" signifies that we pick (K_1, K_2) that minimizes $(1 + 2(K_1 + K_2)/N) \times \bar{L}(K_1, K_2)$. "Multiplicative" signifies that we pick (K_1, K_2) that minimizes $(1 + 2K_1K_2/N) \times \bar{L}(K_1, K_2)$. In this figure, we plot the empirical probability of selecting each pair across $M = 1000$ Monte Carlo samples. The choice set of (K_1, K_2) is the nine pairs put on the horizontal axis.

Figure 8: Share of (K_1, K_2) selected under DGP-NL2 (J-fold cross validation)



DGP-NL2: $T = 0.1(\sum_{j=1}^{2} X_j)^2 + \xi$ and $Y = 1/2 + [(1/2)\sum_{j=1}^{2} X_j]^2 + T + \epsilon$, where $X_1, X_2 \overset{i.i.d.}{\sim} N(0,1)$. K_1 and K_2 are the dimensions of the polynomials of T and $\boldsymbol{X} = (X_1, X_2)^\top$, respectively. We pick (K_1, K_2) that minimizes the loss function of the J-fold cross validation with $J \in \{5, 10\}$. In this figure, we plot the empirical probability of selecting each pair across $M = 1000$ Monte Carlo samples. The choice set of (K_1, K_2) is the nine pairs put on the horizontal axis.

References

Ai, C., O. Linton, K. Motegi, and Z. Zhang (2020): "A Unified Framework for Efficient Estimation of General Treatment Models," Discussion paper, Chinese University of Hong Kong, Shenzhen.

Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. The Johns Hopkins University Press.

Cattaneo, M. D. (2010): "Efficient semiparametric estimation of multi-valued treatment effects under ignorability," *Journal of Econometrics*, 155(2), 138–154.

Chen, X., O. Linton, and I. Van Keilegom (2003): "Estimation of Semiparametric Models When the Criterion Function Is Not Smooth," *Econometrica*, 71(5), 1591–1608.

Firpo, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75(1), 259–276.

Fong, C., C. Hazlett, and K. Imai (2018): "Covariate Balancing Propensity Score for a Continuous Treatment: Application to the Efficacy of Political Advertisements," *Annals of Applied Statistics*, 12(1), 156–177.

Hahn, J. (1998): "On the role of the propensity score in efficient semiparametric estimation of average treatment effects," *Econometrica*, 66(2), 315–331.

Newey, W. K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79(1), 147–168.

Newey, W. K., and D. McFadden (1994): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, chap. 36, pp. 2111–2245. Citeseer.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994): "Estimation of regression coefficients when some regressors are not always observed," *Journal of the American Statistical Association*, 89(427), 846–866.

Tchetgen Tchetgen, E. J., and I. Shpitser (2012): "Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness, and sensitivity analysis," *The Annals of Statistics*, 40(3), 1816–1845.

van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press.