A Simple and Efficient Estimation of Average Treatment Effects in Models With Unmeasured Confounders

Chunrong Ai

Chinese University of Hong Kong, Shenzhen, China

Hubei University of Economics, Wuhan, China

Lukang Huang

Renmin University of China, Beijing, China

Zheng Zhang

Renmin University of China, Beijing, China

Supplementary Material

In this Supplementary Material, we provide additional simulation results and technical proofs for all results presented in the main body of the paper.

S1 Simulation Studies

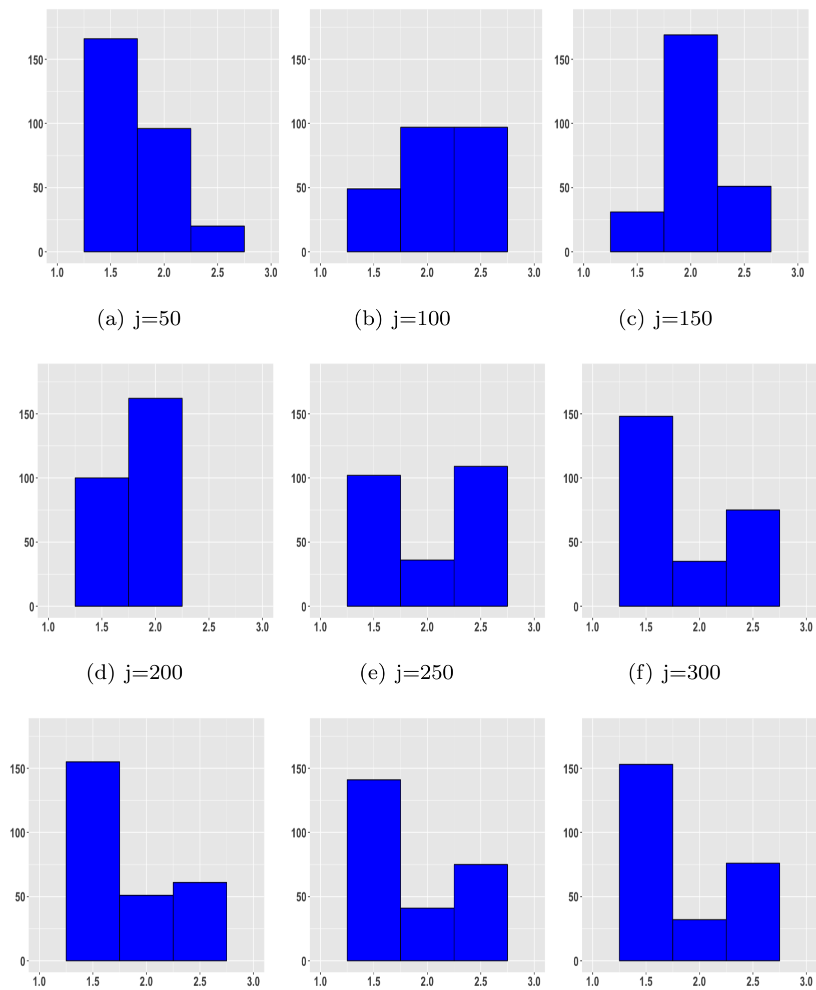
S1.1 Empirical Distributions of Estimated Weighting Functions

At the suggestion of the reviewer, we investigate the stability of the covariate-balanced weights $\{\widehat{w}_{K_1}(1|X_i)\}_{i=1}^N$, $\{\widehat{w}_{K_1}(0|X_i)\}_{i=1}^N$, and $\{\widehat{d}_{K_2}(X_i)\}_{i=1}^N$ with data drive smoothing parameters. Figure 1, 2 and 3 plot their empirical distributions in the j^{th} Monte Carlo run, for $j \in \{50, 100, 150, 200, 250, 300, 350, 400, 450\}$ and $N = 500$. The data generating process (DGP) follows the scenario I in Section 8 of the main text. The plots show that $\{\widehat{w}_{K_1}(1|X_i)\}_{i=1}^N$ are distributed over the interval $[1.25, 2.75]$, $\{\widehat{w}_{K_1}(0|X_i)\}_{i=1}^N$ are distributed over the interval $[1.25, 3.75]$, and $\{\widehat{d}_{K_2}(X_i)\}_{i=1}^N$ are distributed over the interval $[-0.625, -0.125] \cup [0.125, 0.625]$, suggesting that the covariate-balanced approach prevents extreme weights.

S1.2 Simulation Results of Estimated ATE with 2-dimensional Covariates

This scenario adds one more covariate to that in the main paper so that $X = (1, X_1, X_2)$ with $X_2 = U_x^2/4$ and U_x uniformly distributed on $(0, 1)$. The true values of the model parameters are now set at $\alpha = (0.1, 0.5, 3)$, $\beta = (0, -0.5, -0.5)$, $\gamma = (0.1, -0.5, -0.5)$, $\zeta = (0, -1, -1)$, $\eta = (-0.5, 1, 1)$. The average treatment effect now is $\tau = 0.287$.

In each Monte Carlo run, a sample of 500 observations and 1000 observations are generated from the above data generating process. Table 1 reports the bias, standard deviation (Stdev), the root mean square error (RMSE) and the coverage probability (CP) at the nominal size $\alpha = 0.95$. Glancing at the table, the naive estimator is still badly biased in this case. MR has a small bias when $N = 1000$ when some functionals ($\mathcal{M}_1$, $\mathcal{M}_2$)

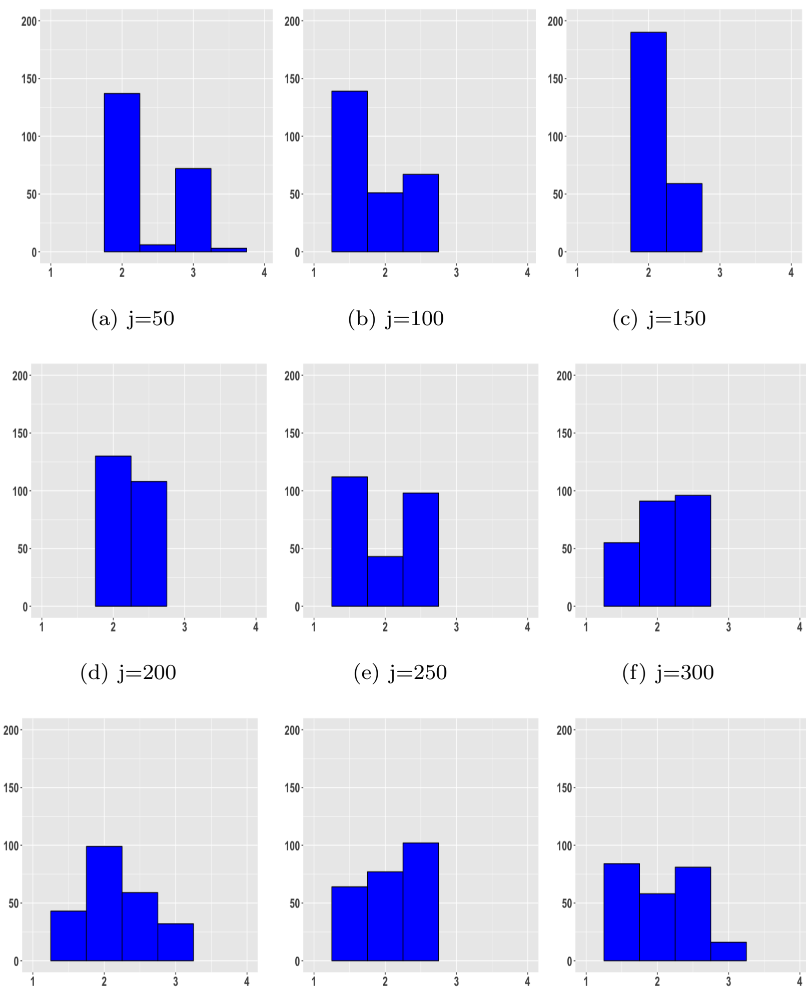


(a) j=50 (b) j=100 (c) j=150

(d) j=200 (e) j=250 (f) j=300

(g) j=350 (h) j=400 (i) j=450

Figure 1: The empirical distribution of $\{\hat{w}_K(1|X_i) : i = 1, \cdots, 500\}$.

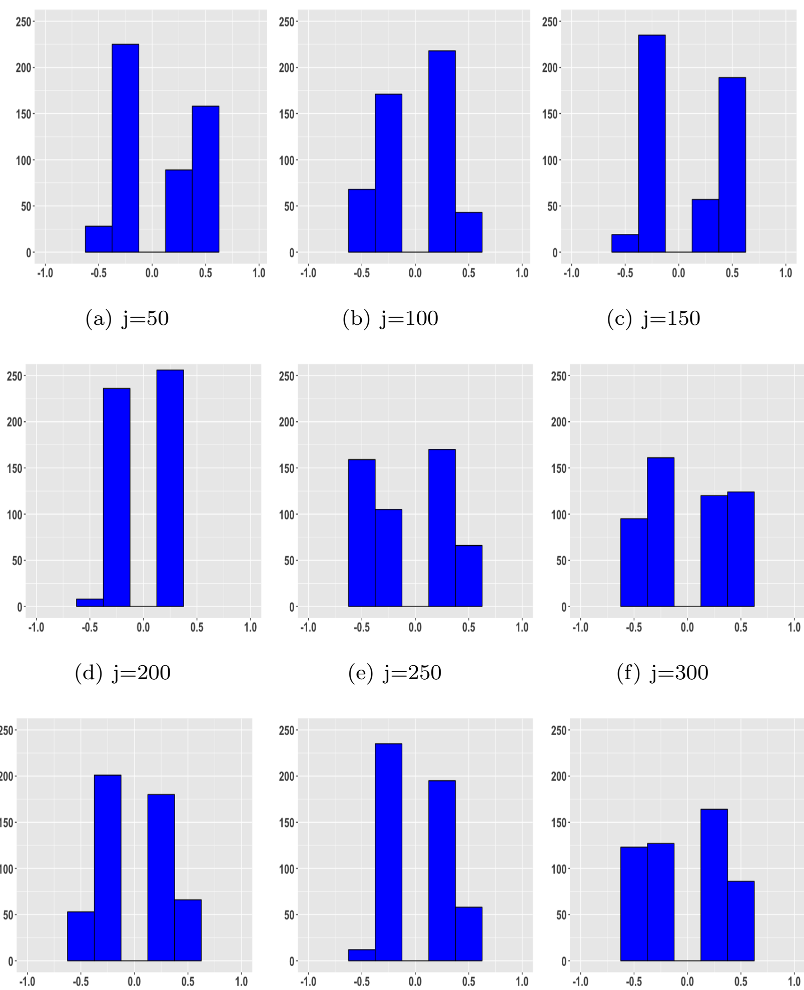


(a) j=50 (b) j=100 (c) j=150

(d) j=200 (e) j=250 (f) j=300

(g) j=350 (h) j=400 (i) j=450

Figure 2: The empirical distribution of $\{\hat{w}_K(0|X_i) : i = 1, \cdots, 500\}$.



(a) j=50 (b) j=100 (c) j=150

(d) j=200 (e) j=250 (f) j=300

(g) j=350 (h) j=400 (i) j=450

Figure 3: The empirical distribution of $\{\hat{d}_K(X_i) : i = 1, \cdots, 500\}$.

or all functionals (*All*) are correctly specified, but when sample size is 500, MR performs badly. B-MR performs well under both sample size when some functionals ($\mathcal{M}_1$, $\mathcal{M}_2$) or all functionals (*All*) are correctly specified, but performs substantially better than MR when most or all functionals are misspecified. The coverage probabilities of both MR and B-MR are close to the nominal size in all cases except when the model is badly misspecified. Despite its out-performance over MR, B-MR is still biased when the model is badly misspecified. Third, HIR produces a large standard deviation when $N = 500$. This due to some estimates $\widehat{E}(D_i|Z_i = 1, X_i) - \widehat{E}(D_i|Z_i = 0, X_i), (i = 1, \cdots, N)$ are very close to zero, see Figure 4 for the empirical distribution. This makes HIR estimator perform badly in the small sample. Fourth, CBE with data-driven smoothing parameters is still unbiased in this case and its coverage probability in all cases is around the nominal size, suggesting that the asymptotic theory is a good approximation.

Table 1: Simulation results of estimated average treatment effects with 2-dimensional covariates

$N = 500$				
Estimators	Bias	Stdev	RMSE	CP
Naive	-0.228	0.045	0.232	0
MR (All)	-0.690	17.22	17.24	0.996
MR ($\mathcal{M}_1$)	0.242	5.259	5.265	0.99
MR ($\mathcal{M}_2$)	-0.327	8.173	8.179	0.994
MR ($\mathcal{M}_3$)	-12.72	403.3	403.5	0.984
MR (None)	60.69	834.709	836.912	0.992
B-MR (All)	0.011	0.151	0.151	0.95
B-MR ($\mathcal{M}_1$)	0.016	0.157	0.158	0.946
B-MR ($\mathcal{M}_2$)	0.041	0.403	0.405	0.952
B-MR ($\mathcal{M}_3$)	0.020	0.154	0.157	0.938
B-MR (None)	0.058	0.462	0.466	0.954
HIR	0.085	2.550	2.551	0.988
CBE	0.003	0.221	0.221	0.956
$N = 1000$				
Estimators	Bias	Stdev	RMSE	CP
Naive	-0.228	0.030	0.230	0
MR (All)	0.001	0.216	0.216	0.976
MR ($\mathcal{M}_1$)	0.007	0.514	0.514	0.984
MR ($\mathcal{M}_2$)	0.027	0.894	0.895	0.980
MR ($\mathcal{M}_3$)	113.9	2355.9	2358.7	0.998
MR (None)	-25.82	450.87	451.61	0.992
B-MR (All)	-0.006	0.104	0.104	0.956
B-MR ($\mathcal{M}_1$)	0.002	0.108	0.108	0.948
B-MR ($\mathcal{M}_2$)	0.068	0.355	0.361	0.938
B-MR($\mathcal{M}_3$)	-0.003	0.105	0.105	0.954
B-MR (None)	0.216	0.443	0.493	0.97
HIR	0.020	0.375	0.376	0.986
CBE	0.009	0.137	0.138	0.952

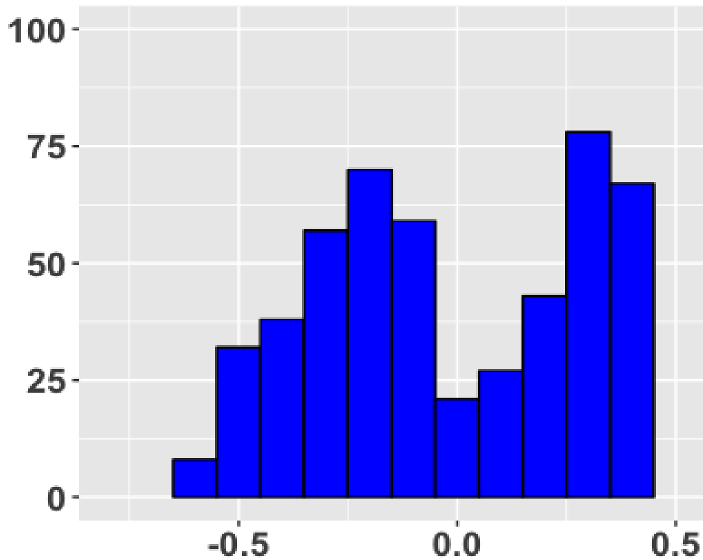


Figure 4: The distribution of $\widehat{E}(D|Z_i = 1, X_i) - \widehat{E}(D|Z_i = 0, X_i)(i = 1, \cdots, 500)$ for 493rd Monte Carlo

S1.3 Simulation Results of Estimated LATE

We consider the following data generating process: $Y = (1-D)\cdot(g(X)+\epsilon)$, $D = ZI(\nu > h(X))$, $Z = I(q(X) > U)$, where the observed covariates X and the unobserved errors ϵ, ν, U are mutually independent with $\epsilon \sim \text{unif}[-1, 1], \nu \sim \text{unif}[-1, 1]$, $U \sim \text{unif}[0, 1]$.

- Scenario I: $X = X_1$, $g(X) = X, h(X) = X_1 - V, q(X) = \tanh(g(X))$ where $X_1 \sim \text{unif}[0, 1], V \sim \text{unif}[0, 1]$. The true value of the local average treatment effects is -0.950.

- Scenario II: $X = (X_1, X_2)$, $g(X) = X_1 + X_2$, $h(X) = X_1 - X_2, q(X) =$

$\Lambda(2 - 1/g(X))$ where $X_i \sim \text{unif}[0,1]$, $(i = 1, 2), \Lambda(x) = e^x/(1 + e^x)$.

The true value of the local average treatment effects is -1.508.

We compute the covariate-balancing estimator (CBE), the inverse probability weighting estimator (IPW) proposed by Donald et al. (2014a), and the inverse probability weighting estimator based on a local polynomial regression (IPW-R) proposed by Donald et al. (2014b). Since Donald et al. (2014b,a) did not discuss how to determine the smoothing parameter in their polynomial approximation. We use the basis $u_2(X) = \{1, X\}$ and $u_3(X) = \{1, X, X^2\}$ respectively in scenario I and $u_3(X) = \{1, X_1, X_2\}$ and $u_6(X) = \{1, X_1, X_2, X_1^2, X_2^2, X_1X_2\}$ respectively in scenario II. The same basis functions are used to compute the CBE estimator so that we can compare the relative performance of all estimators.

Tables 2 and 3 report the bias, standard deviation (Stdev), the root mean square error (RMSE) and the coverage probability (CP) at the nominal size $\alpha = 0.95$. These tables reveal that:

1. In Scenario I, when $u_2(X)$ is used, both IPW and IPW-R are biased and both have incorrect coverage probability. CBE on the other hand is practically unbiased and its coverage probability is close to the nominal size. When $u_3(X)$ is used, the relative performance of the three estimators is the same as when $u_2(X)$ is used.

2. The relative performance of all three estimators in Scenario II is similar to their relative performance in Scenario I.

3. The CBE with data-driven smoothing parameters is practically unbiased and has correct coverage probability in all cases.

Table 2: Simulation results of estimated local average treatment effects for Scenario I

$N = 500$					
Estimators		Bias	Stdev	RMSE	CP
IPW	$u_2(X)$	-0.096	0.055	0.111	0.662
	$u_3(X)$	-0.033	0.039	0.051	0.862
IPW-R	$u_2(X)$	-0.065	0.126	0.142	0.996
	$u_3(X)$	-0.037	0.351	0.353	0.962
CBE	$u_2(X)$	-0.007	0.027	0.028	0.95
	$u_3(X)$	-0.001	0.027	0.027	0.952
CBE	Data-driven	-0.007	0.027	0.028	0.950
$N = 1000$					
Estimators		Bias	Stdev	RMSE	CP
IPW	$u_2(X)$	-0.094	0.043	0.103	0.418
	$u_3(X)$	-0.032	0.028	0.043	0.786
IPW-R	$u_2(X)$	-0.057	0.082	0.100	0.99
	$u_3(X)$	-0.035	0.276	0.278	0.974
CBE	$u_2(X)$	-0.008	0.020	0.021	0.938
	$u_3(X)$	-0.001	0.020	0.020	0.942
CBE	Data-driven	-0.008	0.020	0.021	0.938

Table 3: Simulation results of estimated local average treatment effects for Scenario II

		$N = 500$			
Estimators		Bias	Stdev	RMSE	CP
IPW	$u_3(X)$	-0.168	0.088	0.190	0.534
	$u_6(X)$	-0.049	0.053	0.072	0.85
IPW-R	$u_3(X)$	-0.025	2.076	2.076	0.994
	$u_6(X)$	-0.119	0.104	0.158	0.802
CBE	$u_3(X)$	0.003	0.037	0.037	0.948
	$u_6(X)$	-0.001	0.038	0.038	0.954
CBE	Data-driven	0.003	0.037	0.037	0.948
		$N = 1000$			
Estimators		Bias	Stdev	RMSE	CP
IPW	$u_3(X)$	-0.171	0.064	0.182	0.216
	$u_6(X)$	-0.052	0.037	0.064	0.704
IPW-R	$u_3(X)$	-0.192	4.907	4.911	0.996
	$u_6(X)$	-0.119	0.065	0.126	0.56
CBE	$u_3(X)$	0.004	0.025	0.041	0.942
	$u_6(X)$	-0.005	0.028	0.028	0.946
CBE	Data-driven	0.004	0.025	0.025	0.942

S1.4 Simulation Studies with High-dimensional Covariates

At the suggestion of the reviewer, we investigate the performance of the proposed estimator when the dimension of covariates is high, e.g. 10 covariates. The DGP is specified by adding 8 more covariates to the scenario II in Section 8 of the main text, so that $X = (1, X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10})$ where $X_3 = I(N_1 > 0.05), X_4 = \sin(|N_2|), X_5 = 0.5U_{1x} + 0.5, X_6 = \{0.5 + 0.5U_{2x}\}^{1/2}, X_7 = -0.5U_{3x} - 0.5, X_8 = (3 - 0.5U_{4x})^3/9, X_9 = 0.5 + 0.5U_{5x}, X_{10} = -0.5U_{6x} - 0.5$, and $\{U_{ix}\}_{i=1}^6$ are $i.i.d.$ uniformly distributed on the interval $(0, 1)$, $\{N_i\}_{i=1}^2$ are $i.i.d.$ standard normal random variables. The true values of nuisance parameters are set at $\alpha = (0.1, 0.5, 4, 0, 0.1, 0.9, 0, 0, 0, 0, 0.3)$, $\beta = (0, -0.5, -0.5, 0, 0, 0, 0, 0, 0, 0, 0)$, $\gamma = (0.1, -0.5, -0.5\ , 0, 0, 0, 0, 0, 0, 0, 0)$, $\zeta = (0, -1, -1, 0, 0, 0, 0, 0, 0, 0, 0)$, $\eta = (-0.5, 1, 1, 0, 0, 0.5, 0, 0, 0, 0, 0)$. The true value of the average treatment effect is $\tau = 0.662$.

Table 4: Simulation results of estimated average treatment effects with 10-dimensional covariates

$N = 500$				
Estimator	Bias	Stdev	RMSE	CP
CBE ($K_1 = 11, K_2 = 11$)	0.889	18.38	18.40	0.986
$N = 1000$				
Estimator	Bias	Stdev	RMSE	CP
CBE ($K_1 = 11, K_2 = 11$)	-1.357	31.44	31.47	0.998

Table 4 reports the bias, standard deviation (Stdev), the root mean square error (RMSE) and the coverage probability (CP) at the nominal size 0.95 from the 500 Monte Carlo runs. The proposed estimator has poor performance owing to the curse-of-dimensionality in nonparametric estimation. Estimation of average treatment effects with high-dimensional covariates will be pursued in the future work.

S2 Assumptions

Assumption 1. $\mathbb{E}\left[\frac{1}{\delta^D(X)^2}\right] < \infty$ *and* $\mathbb{E}\left[\frac{Y^2}{\delta^D(X)^4}\right] < \infty$.

Assumption 2. *The support $\mathcal{X}$ of the r-dimensional covariate X is a Cartesian product of r compact intervals.*

Assumption 3. *There exist two positive constants $\overline{C}$ and $\underline{C}$ such that*

$$0 < \underline{C} \leq \lambda_{\min}\left(\mathbb{E}\left[u_K(X)u_K^\top(X)\right]\right) \leq \lambda_{\max}\left(\mathbb{E}\left[u_K(X)u_K^\top(X)\right]\right) \leq \overline{C} < \infty,$$

where $\lambda_{\max}\left(\mathbb{E}\left[u_K(X)u_K^\top(X)\right]\right)$ (resp. $\lambda_{\min}\left(\mathbb{E}\left[u_K(X)u_K^\top(X)\right]\right)$) denotes the largest (resp. smallest) eigenvalue of $\mathbb{E}\left[u_K(X)u_K^\top(X)\right]$.

Assumption 4. *There exist three positive constants $\infty > \eta_1 > \eta_2 > 1 > \eta_3 > 0$ such that $\eta_2 \leq f^{-1}(z|x) \leq \eta_1$ and $-\eta_3 \leq \delta^D(x) \leq \eta_3$, $\forall (z,x) \in \{0,1\} \times \mathcal{X}$.*

Assumption 5. *There exist λ_K, β_K, and γ_K in $\mathbb{R}^K$ and $\alpha > 0$ such that for any $z \in \{0,1\}$,*

$$\sup_{x\in\mathcal{X}}\left|(\rho'^{-1}\left(f^{-1}(z|x)\right) - z\cdot\lambda_K^\top u_K(x) - (1-z)\cdot\beta_K^\top u_K(x)\right| = O(K^{-\alpha}),$$

$$\sup_{x\in\mathcal{X}}\left|(\rho_1'^{-1}\left(\delta^D(x)\right) - \gamma_K^\top u_K(x)\right| = O(K^{-\alpha}).$$

Assumption 6. *$\zeta(K)^4K^3/N \to 0$ and $\sqrt{N}K^{-\alpha} \to 0$, where $\zeta(K) = \sup_{x\in\mathcal{X}} \|u_K(x)\|$ and $\|\cdot\|$ is the usual Frobenius norm defined by $\|A\| = \sqrt{tr(AA^\top)}$ for any matrix A.*

S3 Proof of Theorem 1

We only show that for any integrable function $u(X)$, $\mathbb{E}[Zw(1|X)u(X))] = \mathbb{E}[u(X)]$ holds if and only if $w(1|X) = f^{-1}(1|X)$. The rest can be similarly established. The sufficient part is obvious. We prove the necessary part. Let $u(X) = \exp(a^\top X \cdot i)$ be the test function, where $a \in \mathbb{R}^r$. By assumption,

$$\begin{aligned}&\mathbb{E}\left[\exp(a^\top X \cdot i)\right] = \mathbb{E}\left[Zw(1|X)\exp(a^\top X \cdot i)\right]\\=&\mathbb{E}\left[Z\left\{w(1|X) - f^{-1}(1|X)\right\}\exp(a^\top X \cdot i)\right] + \mathbb{E}\left[Zf^{-1}(1|X)\exp(a^\top X \cdot i)\right].\end{aligned}$$

By the tower property of conditional expectation, $\mathbb{E}\left[Zf^{-1}(1|X)\exp(a^\top X \cdot i)\right] = \mathbb{E}\left[\exp(a^\top X \cdot i)\right]$. Then

$$0 = \mathbb{E}\left[Z(w(1|X) - f^{-1}(1|X))\exp(a^\top X \cdot i)\right] = \mathbb{E}[\{f(1|X)w(1|X) - 1\}\exp(a^\top X \cdot i)]$$

holds for all $a \in \mathbb{R}^r$. Due to the uniqueness of Fourier transform we can obtain $f(1|X)w(1|X) - 1 = 0$, which implies $w(1|X) = f^{-1}(1|X)$.

S4 Convergence Rates of $\hat{w}_K(1|X)$, $\hat{w}_K(0|X)$ and $\hat{d}_K(X)$

Proposition 1. *Under Assumptions 1-6, we have*

$$\sup_{x\in\mathcal{X}} |\hat{w}_K(1|x) - f(1|x)^{-1}| = O_p\left(\zeta(K)K^{-\alpha} + \zeta(K)\sqrt{\frac{K}{N}}\right) ,$$
$$\int_{\mathcal{X}} |\hat{w}_K(1|x) - f(1|x)^{-1}|^2 dF_X(x) = O_p\left(K^{-2\alpha} + \frac{K}{N}\right) ,$$
$$\frac{1}{N}\sum_{i=1}^N |\hat{w}_K(1|X_i) - f(1|X_i)^{-1}|^2 = O_p\left(K^{-2\alpha} + \frac{K}{N}\right) ,$$

and

$$\sup_{x\in\mathcal{X}} |\hat{w}_K(0|x) - f(0|x)^{-1}| = O_p\left(\zeta(K)K^{-\alpha} + \zeta(K)\sqrt{\frac{K}{N}}\right) ,$$
$$\int_{\mathcal{X}} |\hat{w}_K(0|x) - f(0|x)^{-1}|^2 dF_X(x) = O_p\left(K^{-2\alpha} + \frac{K}{N}\right) ,$$
$$\frac{1}{N}\sum_{i=1}^N |\hat{w}_K(0|X_i) - f(0|X_i)^{-1}|^2 = O_p\left(K^{-2\alpha} + \frac{K}{N}\right) ,$$

and

$$\sup_{x\in\mathcal{X}} |\hat{d}_K(x) - \delta^D(x)| = O_p\left(\zeta(K)K^{-\alpha} + \zeta(K)\sqrt{\frac{K}{N}}\right) ,$$
$$\int_{\mathcal{X}} |\hat{d}_K(x) - \delta^D(x)|^2 dF_X(x) = O_p\left(K^{-2\alpha} + \frac{K}{N}\right) ,$$
$$\frac{1}{N}\sum_{i=1}^N |\hat{d}_K(X_i) - \delta^D(X_i)|^2 = O_p\left(K^{-2\alpha} + \frac{K}{N}\right) .$$

Proposition 1 provides the convergence rates of $\hat{w}_K(1|X) \to f^{-1}(1|X)$, $\hat{w}_K(0|X) \to f^{-1}(0|X)$ and $\hat{d}_K(X) \to \delta^D(X)$. It directly follows from Lemmas 1 and 2 stated in the following subsections. We introduce some notations which will be used later. By Assumption 3, we assume the sieve basis $u_K(X)$ is orthonormalized, namely

$$\mathbb{E}[u_K(X)u_K^\top(X)] = I_K. \tag{S4.1}$$

Since

$$(\hat{\lambda}_K, \hat{\beta}_K) = \arg\max_{\lambda,\beta} \hat{G}(\lambda, \beta).$$

It is obvious to see that $\hat{\lambda}_K = \arg\max \hat{G}_1(\lambda)$ and $\hat{\beta}_K = \arg\max \hat{G}_2(\beta)$, where

$$\hat{G}_1(\lambda) = \frac{1}{N}\sum_{i=1}^N \left\{Z_i\rho'(\lambda^\top u_K(X_i)) - \lambda^\top u_K(X_i)\right\},$$
$$\hat{G}_2(\beta) = \frac{1}{N}\sum_{i=1}^N \left\{(1-Z_i)\rho'(\beta^\top u_K(X_i)) - \beta^\top u_K(X_i)\right\}.$$

Let $G_1^*(\lambda)$, $G_2^*(\beta)$, $H^*(\gamma)$, λ_K^*, β_K^*, γ_K^*, $w_K^*(1|X)$, $w_K^*(0|X)$ and $d_K^*(X)$ be the theoretical counterparts of $\hat{G}_1(\lambda)$, $\hat{G}_2(\beta)$, $\hat{H}(\gamma)$, $\hat{\lambda}_K$, $\hat{\beta}_K$, $\hat{\gamma}_K$, $\hat{w}_K(1|X)$,

$\hat{w}_K(0|X)$ and $\hat{d}_K(X)$ respectively:

$$
\begin{aligned}
&G_1^*(\lambda) := \mathbb{E}[\hat{G}_1(\lambda)] = \mathbb{E}\left[Z\rho\left(\lambda^\top u_K(X)\right)\right] - \mathbb{E}[\lambda^\top u_K(X)],\\
&G_2^*(\beta) := \mathbb{E}[\hat{G}_2(\beta)] = \mathbb{E}\left[(1-Z)\rho\left(\beta^\top u_K(X)\right)\right] - \mathbb{E}[\beta^\top u_K(X)],\\
&H^*(\gamma) := \mathbb{E}\left[\rho_1\left(\gamma^\top u_K(X)\right)\right] - \mathbb{E}\left[D\left\{\frac{Z}{f(1|X)} - \frac{1-Z}{f(0|X)}\right\}\gamma^\top u_K(X)\right],\\
&\lambda_K^* := \arg\max_{\lambda\in\mathbb{R}^K} G_1^*(\lambda)\ ,\ \beta_K^* := \arg\max_{\beta\in\mathbb{R}^K} G_2^*(\lambda),\ \gamma_K^* := \arg\max_{\gamma\in\mathbb{R}^K} H^*(\gamma)\ ,\\
&w_K^*(1|X) := \rho'\left((\lambda_K^*)^\top u_K(X)\right),\ w_K^*(0|X) := \rho'\left((\beta_K^*)^\top u_K(X)\right),\\
&d_K^*(X) := \rho_1'((\gamma_K^*)^\top u_K(X)).
\end{aligned}
$$

S4.1 Lemma 1

The first lemma gives the approximation rate of the intermediate quantities $w_K^*(1|x)$, $w_K^*(0|x)$ and $d_K^*(x)$. Recall the notation $\zeta(K) = \sup_{x\in\mathcal{X}} \|u_K(x)\|$.

Lemma 1. *Under Assumptions 1-6, we have*

$$
\begin{aligned}
&\sup_{x\in\mathcal{X}}\left|f(1|x)^{-1} - w_K^*(1|x)\right| = O\left(\zeta(K)K^{-\alpha}\right)\ ,\\
&\int_{\mathcal{X}}\left|f(1|x)^{-1} - w_K^*(1|x)\right|^2 dF_X(x) = O\left(K^{-2\alpha}\right)\ ,\\
&\frac{1}{N}\sum_{i=1}^N\left|f(1|X_i)^{-1} - w_K^*(1|X_i)\right|^2 = O_p\left(K^{-2\alpha}\right)\ ,
\end{aligned}
$$

and

$$\sup_{x\in\mathcal{X}}\left|f(0|x)^{-1}-w_K^*(0|x)\right| = O\left(\zeta(K)K^{-\alpha}\right) \ ,$$
$$\int_{\mathcal{X}}\left|f(0|x)^{-1}-w_K^*(0|x)\right|^2 dF_X(x) = O\left(K^{-2\alpha}\right) \ ,$$
$$\frac{1}{N}\sum_{i=1}^N\left|f(0|X_i)^{-1}-w_K^*(0|X_i)\right|^2 = O_p\left(K^{-2\alpha}\right) \ ,$$

and

$$\sup_{x\in\mathcal{X}}\left|\delta^D(x)-d_K^*(x)\right| = O\left(\zeta(K)K^{-\alpha}\right) \ ,$$
$$\int_{\mathcal{X}}\left|\delta^D(x)-d_K^*(x)\right|^2 dF_X(x) = O\left(K^{-2\alpha}\right) \ ,$$
$$\frac{1}{N}\sum_{i=1}^N\left|\delta^D(X_i)-d_K^*(X_i)\right|^2 = O_p\left(K^{-2\alpha}\right) \ .$$

Proof. We only prove

$$\sup_{x\in\mathcal{X}}\left|f(1|x)^{-1}-w_K^*(1|x)\right| = O\left(\zeta(K)K^{-\alpha}\right) \ ,$$
$$\int_{\mathcal{X}}\left|f(1|x)^{-1}-w_K^*(1|x)\right|^2 dF_X(x) = O\left(K^{-2\alpha}\right) \ ,$$
$$\frac{1}{N}\sum_{i=1}^N\left|f(1|X_i)^{-1}-w_K^*(1|X_i)\right|^2 = O_p\left(K^{-2\alpha}\right) \ ,$$

and similar argument can be applied to obtain the other claims.

By Assumption 4, $f(1|x)^{-1} \in [\eta_2, \eta_1]$, $\forall x \in \mathcal{X}$ and the fact $(\rho')^{-1}$ is strictly decreasing, we have

$$(\rho')^{-1}(\eta_1) \leq \inf_{x\in\mathcal{X}} (\rho')^{-1}\left(\frac{1}{f(1|x)}\right) \leq \sup_{x\in\mathcal{X}} (\rho')^{-1}\left(\frac{1}{f(1|x)}\right) \leq (\rho')^{-1}(\eta_2)\ .$$

By Assumption 5,

$$\sup_{x\in\mathcal{X}} \left|(\rho')^{-1}\left(\frac{1}{f(1|x)}\right) - \lambda_K^\top u_K(x)\right| < C_1 K^{-\alpha}, \tag{S4.2}$$

where $C_1 > 0$ is a universal constant. Then we have

$$\lambda_K^\top u_K(x) \in \left((\rho')^{-1}\left(\frac{1}{f(1|x)}\right) - C_1K^{-\alpha}, (\rho')^{-1}\left(\frac{1}{f(1|x)}\right) + C_1K^{-\alpha}\right) \tag{S4.3}$$

$$\subset \left[(\rho')^{-1}(\eta_1) - C_1K^{-\alpha}, (\rho')^{-1}(\eta_2) + C_1K^{-\alpha}\right]\ ,\ \forall x \in \mathcal{X}\ ,$$

and

$$\rho'(\lambda_K^\top u_K(x) + C_1K^{-\alpha}) - \rho'(\lambda_K^\top u_K(x)) < \frac{1}{f(1|x)} - \rho'(\lambda_K^\top u_K(x))$$
$$< \rho'(\lambda_K^\top u_K(x) - C_1K^{-\alpha}) - \rho'(\lambda_K^\top u_K(x))\ ,\ \forall x \in \mathcal{X}.$$

By Mean Value Theorem, for large enough K, there exists

$$\xi_1(x) \in (\lambda_K^\top u_K(x), \lambda_K^\top u_K(x) + C_1K^{-\alpha}) \subset [(\rho')^{-1}(\eta_1) - C_1K^{-\alpha}, (\rho')^{-1}(\eta_2) + 2C_1K^{-\alpha}] \subset \Gamma_1$$

$$\xi_2(x) \in (\lambda_K^\top u_K(x) - C_1 K^{-\alpha}, \lambda_K^\top u_K(x)) \subset [(\rho')^{-1}(\eta_1) - 2C_1K^{-\alpha}, (\rho')^{-1}(\eta_2) + C_1K^{-\alpha}] \subset \Gamma_1 \ ,$$

where

$$\Gamma_1 := \left[(\rho')^{-1}(\eta_1) - 1, (\rho')^{-1}(\eta_2) + 1\right] \ ,$$

such that

$$\rho'(\lambda_K^\top u_K(x) + C_1K^{-\alpha}) - \rho'(\lambda_K^\top u_K(x)) = \rho''(\xi_1(x))C_1K^{-\alpha} \geq \inf_{y\in\Gamma_1} \rho''(y)C_1K^{-\alpha}$$

$$\rho'(\lambda_K^\top u_K(x) - C_1K^{-\alpha}) - \rho'(\lambda_K^\top u_K(x)) = -\rho''(\xi_2(x))C_1K^{-\alpha} \leq \sup_{y\in\Gamma_1} -\rho''(y)C_1K^{-\alpha} \ .$$

Let $a := \max\left\{-\inf_{y\in\Gamma_1}\rho''(y), \sup_{y\in\Gamma_1} -\rho''(y)\right\}$, which is a finite positive constant because the set Γ_1 is compact and the function $\rho''(y)$ is continuous. Therefore, we have

$$\sup_{x\in\mathcal{X}}\left|\frac{1}{f(1|x)} - \rho'(\lambda_K^\top u_K(x))\right| < aC_1K^{-\alpha} \ . \tag{S4.4}$$

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

$$\Lambda_K := \left\{\lambda \in \mathbb{R}^K : \|\lambda - \lambda_K\| \leq C_2K^{-\alpha}\right\}.$$

For sufficiently large K, by (S4.3), Assumption 6, we have that $\forall \lambda \in \Lambda_K$,

$\forall x \in \mathcal{X}$,

$$
\begin{aligned}
&|\lambda^\top u_K(x) - \lambda_K^\top u_K(x)| = |(\lambda - \lambda_K)^\top u_K(x)| \leq \|\lambda - \lambda_K\| \|u_K(x)\| \leq C_2 K^{-\alpha} \zeta(K) \\
\Rightarrow & \lambda^\top u_K(x) \in (\lambda_K^\top u_K(x) - C_2 K^{-\alpha}\zeta(K), \lambda_K^\top u_K(x) + C_2 K^{-\alpha}\zeta(K)) \\
& \qquad \subset \left[(\rho')^{-1}(\eta_1) - C_1 K^{-\alpha} - C_2 K^{-\alpha}\zeta(K), (\rho')^{-1}(\eta_2) + C_1 K^{-\alpha} + C_2 K^{-\alpha}\zeta(K)\right] \\
& \qquad \subset \Gamma_1 \ . \qquad \text{(S4.5)}
\end{aligned}
$$

By (S4.4), (S4.5), and the fact $\mathbb{E}[u_K(X) u_K(X)^\top] = I_K$, we can deduce that

$$
\begin{aligned}
&\left\|(G_1^*)'(\lambda_K)\right\| = \left\|\mathbb{E}\left[f(1|X)\rho'\left(\lambda_K^\top u_K(X)\right) u_K(X) - u_K(X)\right]\right\| \\
=& \left\|\mathbb{E}\left[f(1|X)\left\{\rho'\left(\lambda_K^\top u_K(X)\right) - \frac{1}{f(1|X)}\right\} u_K(X)\right]\right\| \\
=& \sqrt{\mathbb{E}\left[\left\{\mathbb{E}\left[f(1|X)\left\{\rho'\left(\lambda_K^\top u_K(X)\right) - \frac{1}{f(1|X)}\right\} u_K(X)^\top\right] \cdot \mathbb{E}\left[u_K(X) u_K(X)^\top\right]^{-1} u_K(X)\right\}^2\right]} \\
=& \left\|\mathbb{E}\left[f(1|X)\left\{\rho'\left(\lambda_K^\top u_K(X)\right) - \frac{1}{f(1|X)}\right\} u_K(X)^\top\right] \cdot \mathbb{E}\left[u_K(X) u_K(X)^\top\right]^{-1} u_K(X)\right\|_{L^2} \\
\leq& \left\|f(1|X)\left\{\rho'\left(\lambda_K^\top u_K(X)\right) - \frac{1}{f(1|X)}\right\}\right\|_{L^2} \\
\leq& \left\|\rho'\left(\lambda_K^\top u_K(X)\right) - \frac{1}{f(1|X)}\right\|_{L^2} \leq a C_1 K^{-\alpha} \ , \qquad \text{(S4.6)}
\end{aligned}
$$

where the third equality follows from the definition of Frobenius norm and the fact $\mathbb{E}[u_K(X) u_K(X)^\top] = I_K$; the first inequality follows from the fact that

$$
\mathbb{E}\left[f(1|X)\left\{\rho'\left(\lambda_K^\top u_K(X)\right) - \frac{1}{f(1|X)}\right\} u_K(X)^\top\right] \cdot \mathbb{E}\left[u_K(X) u_K(X)^\top\right]^{-1} u_K(X)
$$

is the $L^2(dF_X)$-projection of $f(1|X)\left\{\rho'\left(\lambda_K^\top u_K(X)\right) - \frac{1}{f(1|X)}\right\}$ on the space spanned by $u_K(X)$.

Then for any $\lambda \in \partial\Lambda_K$, i.e. $\|\lambda - \lambda_K\| = C_2 K^{-\alpha}$, by Mean Value Theorem we can deduce that

$$
\begin{aligned}
& G_1^*(\lambda) - G_1^*(\lambda_K) \\
&= (\lambda - \lambda_K)^\top (G_1^*)'(\lambda_K) + \frac{1}{2}(\lambda - \lambda_K)^\top (G_1^*)''(\bar{\lambda}_K)(\lambda - \lambda_K) \\
&\le \|\lambda - \lambda_K\| \|(G_1^*)'(\lambda_K)\| + \frac{1}{2}(\lambda - \lambda_K)^\top \mathbb{E}\left[f(1|X)\rho''\left(\bar{\lambda}_K^\top u_K(X)\right) u_K(X) u_K^\top(X)\right](\lambda - \lambda_K) \\
&\le \|\lambda - \lambda_K\| \|(G_1^*)'(\lambda_K)\| - \frac{a_1}{2} \cdot \frac{1}{\eta_1}(\lambda - \lambda_K)^\top \mathbb{E}[u_K(X) u_K^\top(X)](\lambda - \lambda_K) \\
&= \|\lambda - \lambda_K^*\| \|(G_1^*)'(\lambda_K)\| - \frac{a_1}{2} \cdot \frac{1}{\eta_1}\|\lambda - \lambda_K\|^2 \\
&= \|\lambda - \lambda_K\| \left(\|(G_1^*)'(\lambda_K)\| - \frac{a_1}{2\eta_1}\|\lambda - \lambda_K\| \right) \\
&\le \|\lambda - \lambda_K\| \left(a C_1 K^{-\alpha} - \frac{a_1}{2\eta_1} \cdot C_2 K^{-\alpha} \right), \qquad \text{(S4.7)}
\end{aligned}
$$

where $a_1 = \inf_{y\in\Gamma_1}\{-\rho''(y)\} > 0$ is a finite positive constant, and the last inequality follows from (S4.6). By choosing

$$C_2 > \frac{2aC_1\eta_1}{a_1} \ ,$$

we can obtain that

$$G_1^*(\lambda) < G_1^*(\lambda_K) \ , \quad \lambda \in \partial\Lambda_K \ .$$

In light of the continuity of G_1^*, there is a local maximum of G_1^* in the interior of Λ_K. On the other hand, G_1^* is a strictly concave function with a unique global maximum point λ_K^*, therefore we can claim

$$\lambda_K^* \in \Lambda_K^\circ \ , \quad i.e. \ \ \|\lambda_K^* - \lambda_K\| \leq C_2 K^{-\alpha} \ . \tag{S4.8}$$

By Mean Value Theorem, for large enough K, there exists $\xi^*(x)$ lying between $(\lambda_K^*)^\top u_K(x)$ and $\lambda_K^\top u_K(x)$, which implies $\xi^*(x) \in \Gamma_1$, such that for any $x \in \mathcal{X}$,

$$\begin{aligned} & |\rho'(\lambda_K^\top u_K(x)) - \rho'((\lambda_K^*)^\top u_K(x))| \qquad \text{(S4.9)} \\ = & |\rho''(\xi^*(x))||\lambda_K^\top u_K(x) - (\lambda_K^*)^\top u_K(x)| \\ \leq & -\rho''(\xi^*(x))\|\lambda_K - \lambda_K^*\|\|u_K(x)\| \leq a_2 C_2 K^{-\alpha}\zeta(K) \ , \end{aligned}$$

where $a_2 = \sup_{\gamma\in\Gamma_1} -\rho''(\gamma) < \infty$. Therefore,

$$\begin{aligned} & \sup_{x\in\mathcal{X}} \left| \frac{1}{f(1|x)} - \rho'((\lambda_K^*)^\top u_K(x)) \right| \qquad \text{(S4.10)} \\ = & \sup_{x\in\mathcal{X}} \left| \frac{1}{f(1|x)} - \rho'(\lambda_K^\top u_K(x)) + \rho'(\lambda_K^\top u_K(x)) - \rho'((\lambda_K^*)^\top u_K(x)) \right| \\ \leq & \sup_{x\in\mathcal{X}} \left| \frac{1}{f(1|x)} - \rho'(\lambda_K^\top u_K(x)) \right| + \sup_{x\in\mathcal{X}} \left| \rho'(\lambda_K^\top u_K(x)) - \rho'((\lambda_K^*)^\top u_K(x)) \right| \\ \leq & a C_1 K^{-\alpha} + a_2 C_2 K^{-\alpha}\zeta(K) \\ \leq & (aC_1 + a_2 C_2) K^{-\alpha}\zeta(K) = O\left(\zeta(K) K^{-\alpha}\right) \ , \end{aligned}$$

where the second inequality follows from (S4.4), (S4.5) and (S4.9).

Similarly, by (S4.4), (S4.5), (S4.8) and the fact $\mathbb{E}[u_K(X)u_K(X)^\top] = I_K$, we can deduce that

$$
\begin{aligned}
&\int_{\mathcal{X}} \left|f(1|x)^{-1} - w_K^*(1|x)\right|^2 dF_X(x) \\
\leq & 2\int_{\mathcal{X}} \left|f(1|x)^{-1} - \rho'\left(\lambda_K^\top u_K(x)\right)\right|^2 dF_X(x) + 2\int_{\mathcal{X}} \left|\rho'\left(\lambda_K^\top u_K(x)\right) - \rho'\left((\lambda_K^*)^\top u_K(x)\right)\right|^2 dF_X(x) \\
\leq & 2\sup_{x\in\mathcal{X}} \left|f(1|x)^{-1} - \rho'\left(\lambda_K^\top u_K(x)\right)\right|^2 + 2\cdot\int_{\mathcal{X}} |\rho''(\xi^*(x))|^2 \cdot |(\lambda_K - \lambda_K^*)^\top u_K(x)|^2 dF_X(x) \\
\leq & 2\sup_{x\in\mathcal{X}} \left|f(1|x)^{-1} - \rho'\left(\lambda_K^\top u_K(x)\right)\right|^2 + 2\sup_{x\in\mathcal{X}} |\rho''(\xi^*(x))|^2 \cdot (\lambda_K - \lambda_K^*)^\top \int_{\mathcal{X}} u_K(x)u_K(x)^\top dF_X(x)\cdot(\lambda_K - \lambda_K^*) \\
= & 2\sup_{x\in\mathcal{X}} \left|f(1|x)^{-1} - \rho'\left(\lambda_K^\top u_K(x)\right)\right|^2 + 2\sup_{x\in\mathcal{X}} |\rho''(\xi^*(x))|^2 \cdot \|\lambda_K - \lambda_K^*\|^2 \\
= & O(K^{-2\alpha}) + O(1)\cdot O(K^{-2\alpha}) = O(K^{-2\alpha})\ .
\end{aligned}
$$

We can also obtain

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N \left|f(1|X_i)^{-1} - w_K^*(1|X_i)\right|^2 \\
\leq & \frac{2}{N}\sum_{i=1}^N \left|f(1|X_i)^{-1} - \rho'\left(\lambda_K^\top u_K(X_i)\right)\right|^2 + \frac{2}{N}\sum_{i=1}^N \left|\rho'\left(\lambda_K^\top u_K(x)\right) - \rho'\left((\lambda_K^*)^\top u_K(X_i)\right)\right|^2 \\
= & \frac{2}{N}\sum_{i=1}^N \left|f(1|X_i)^{-1} - \rho'\left(\lambda_K^\top u_K(X_i)\right)\right|^2 + \frac{2}{N}\sum_{i=1}^N \left|\rho''\left(\xi^*(x)\right)^\top (\lambda_K - \lambda_K^*) u_K(X_i)\right|^2 \\
\leq & 2\sup_{x\in\mathcal{X}} \left|f(1|x)^{-1} - \rho'\left(\lambda_K^\top u_K(x)\right)\right|^2 + 2\sup_{x\in\mathcal{X}} |\rho''(\xi^*(x))|^2 \cdot (\lambda_K - \lambda_K^*)^\top \left\{\frac{1}{N}\sum_{i=1}^N u_K(X_i)u_K(X_i)^\top\right\}(\lambda_K - \lambda_K^*) \\
\leq & 2\sup_{x\in\mathcal{X}} \left|f(1|x)^{-1} - \rho'\left(\lambda_K^\top u_K(x)\right)\right|^2 + 2\sup_{x\in\mathcal{X}} |\rho''(\xi^*(x))|^2 \cdot \lambda_{\max}\left\{\frac{1}{N}\sum_{i=1}^N u_K(X_i)u_K(X_i)^\top\right\}\|\lambda_K - \lambda_K^*\|^2 \\
= & O(K^{-2\alpha}) + O(1)\cdot O_p(1)\cdot O(K^{-2\alpha}) = O_p(K^{-2\alpha})\ ,
\end{aligned}
$$

where $\lambda_{\max}(A)$ denotes the largest eigenvalue of a matrix A; the second equality follows from Chebyshev's inequality and the following facts

$$
\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^N u_K(X_i)u_K(X_i)^\top - \mathbb{E}[u_K(X)u_K(X)^\top]\right\|^2\right] \tag{S4.11}
$$

$$
\begin{aligned}
&= \frac{1}{N}\mathbb{E}\left[\left\| u_K(X)u_K(X)^\top - \mathbb{E}[u_K(X)u_K(X)^\top] \right\|^2\right] \\
&= \frac{1}{N}\mathbb{E}\left[\mathrm{tr}\left\{ u_K(X)u_K(X)^\top u_K(X)u_K(X)^\top \right\}\right] - \frac{1}{N}\mathrm{tr}\left\{\mathbb{E}[u_K(X)u_K(X)^\top]\cdot \mathbb{E}[u_K(X)u_K(X)^\top]\right\} \\
&\leq \frac{1}{N}\cdot \zeta(K)^2\mathbb{E}\left[\|u_K(X)\|^2\right] = \zeta(K)^2\frac{K}{N} \to 0 \ ,
\end{aligned}
$$

and $\mathbb{E}\left[u_K(X)u_K(X)^\top\right] = I_K$. □

S4.2 Lemma 2

Lemma 2. *Under Assumptions 1-6, we have*

$$
\begin{aligned}
&\|\hat{\lambda}_K - \lambda_K^*\| = O_p\left(\sqrt{\frac{K}{N}}\right) \ , \\
&\sup_{x\in\mathcal{X}} |\hat{w}_K(1|x) - w_K^*(1|x)| = O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right) \ , \\
&\int_{\mathcal{X}} |\hat{w}_K(1|x) - w_K^*(1|x)|^2 dF_X(x) = O_p\left(\frac{K}{N}\right) \ , \\
&\frac{1}{N}\sum_{i=1}^{N} |\hat{w}_K(1|X_i) - w_K^*(1|X_i)|^2 = O_p\left(\frac{K}{N}\right) \ ,
\end{aligned}
$$

and

$$
\begin{aligned}
&\|\hat{\beta}_K - \beta_K^*\| = O_p\left(\sqrt{\frac{K}{N}}\right) \ , \\
&\sup_{x\in\mathcal{X}} |\hat{w}_K(0|x) - w_K^*(0|x)| = O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right) \ , \\
&\int_{\mathcal{X}} |\hat{w}_K(0|x) - w_K^*(0|x)|^2 dF_X(x) = O_p\left(\frac{K}{N}\right) \ ,
\end{aligned}
$$

$$\frac{1}{N}\sum_{i=1}^{N}|\hat{w}_K(0|X_i) - w_K^*(0|X_i)|^2 = O_p\left(\frac{K}{N}\right) \ ,$$

and

$$\begin{aligned}
&\|\hat{\gamma}_K - \gamma_K^*\| = O_p\left(K^{-\alpha} + \sqrt{\frac{K}{N}}\right) \ , \\
&\sup_{x\in\mathcal{X}}|\hat{d}_K(x) - d_K^*(x)| = O_p\left(\zeta(K)K^{-\alpha} + \zeta(K)\sqrt{\frac{K}{N}}\right) \ , \\
&\int_{\mathcal{X}}|\hat{d}_K(x) - d_K^*(x)|^2 dF_X(x) = O_p\left(K^{-2\alpha} + \frac{K}{N}\right) \ , \\
&\frac{1}{N}\sum_{i=1}^{N}|\hat{d}_K(X_i) - d_K^*(X_i)|^2 = O_p\left(K^{-2\alpha} + \frac{K}{N}\right) \ .
\end{aligned}$$

Proof. We first prove

$$\begin{aligned}
&\|\hat{\lambda}_K - \lambda_K^*\| = O_p\left(\sqrt{\frac{K}{N}}\right) \ , \\
&\sup_{x\in\mathcal{X}}|\hat{w}_K(1|x) - w_K^*(1|x)| = O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right) \ , \\
&\int_{\mathcal{X}}|\hat{w}_K(1|x) - w_K^*(1|x)|^2 dF_X(x) = O_p\left(\frac{K}{N}\right) \ , \\
&\frac{1}{N}\sum_{i=1}^{N}|\hat{w}_K(1|X_i) - w_K^*(1|X_i)|^2 = O_p\left(\frac{K}{N}\right) \ .
\end{aligned}$$

Define

$$\hat{S}_N := \frac{1}{N}\sum_{i=1}^{N} Z_i u_K(X_i) u_K(X_i)^\top \ ,$$

obviously $\hat{S}_N$ is a symmetric matrix and $\mathbb{E}[\hat{S}_N] = \mathbb{E}[f(1|X)u_K(X)u_K^\top(X)]$. We have

$$\begin{aligned}
&\mathbb{E}\left[\left\|\hat{S}_N - \mathbb{E}\left[f(1|X)u_K(X)u_K^\top(X)\right]\right\|^2\right] \\
&= \mathrm{tr}\left(\mathbb{E}[\hat{S}_N\hat{S}_N] - 2\mathbb{E}[\hat{S}_N]\mathbb{E}[f(1|X)u_K(X)u_K^\top(X)] + \mathbb{E}[f(1|X)u_K(X)u_K^\top(X)]\mathbb{E}[f(1|X)u_K(X)u_K^\top(X)]\right) \\
&= \mathrm{tr}\Bigg(\mathbb{E}\left[\frac{1}{N^2}\sum_{i=1}^N Z_i^2 u_K(X_i)u_K^\top(X_i)u_K(X_i)u_K^\top(X_i)\right] + \mathbb{E}\left[\frac{1}{N^2}\sum_{i,j=1,i\neq j}^N Z_iZ_j u_K(X_i)u_K^\top(X_i)u_K(X_j)u_K^\top(X_j)\right] \\
&\quad - \mathbb{E}[f(1|X)u_K(X)u_K^\top(X)]\mathbb{E}[f(1|X)u_K(X)u_K^\top(X)]\Bigg) \\
&= \frac{1}{N}\mathbb{E}[f(1|X)u_K(X)^\top u_K(X)u_K(X)^\top u_K(X)] - \frac{1}{N}\mathrm{tr}\left(\mathbb{E}[f(1|X)u_K(X)u_K^\top(X)]\cdot\mathbb{E}[f(1|X)u_K(X)u_K^\top(X)]\right) \\
&\leq \zeta(K)^2\frac{K}{N}\ ,
\end{aligned} \tag{S4.12}$$

where the last inequality follows from the facts $\sup_{x\in\mathcal{X}}\|u_K(x)\| \leq \zeta(K)$, $0 < f(1|x) < 1$ and $\mathbb{E}[u_K^\top(X)u_K(X)] = K$.

Consider the event set

$$E_{N,\eta_1} := \left\{(\lambda-\lambda_K^*)^\top\hat{S}_N(\lambda-\lambda_K^*) > (\lambda-\lambda_K^*)^\top\left(\mathbb{E}[f(1|X)u_K(X)u_K^\top(X)] - \frac{1}{2\eta_1}I_K\right)(\lambda-\lambda_K^*), \lambda\neq\lambda_K^*\right\}\ .$$

By Chebyshev's inequality and (S4.12) and Assumption 6 we have

$$\begin{aligned}
&\mathbb{P}\left(\left|(\lambda-\lambda_K^*)^\top\hat{S}_N(\lambda-\lambda_K^*) - (\lambda-\lambda_K^*)^\top\mathbb{E}[f(1|X)u_K(X)u_K^\top(X)](\lambda-\lambda_K^*)\right| \geq \frac{1}{2\eta_1}\|\lambda-\lambda_K^*\|^2, \lambda\neq\lambda_K^*\right) \\
&\leq \frac{4\eta_1^2\|\lambda-\lambda_K^*\|^4\mathbb{E}\left[\left\|\hat{S}_N - \mathbb{E}\left[f(1|X)u_K(X)u_K^\top(X)\right]\right\|^2\right]}{\|\lambda-\lambda_K^*\|^4} \leq \frac{4\eta_1^2\zeta(K)^2K}{N} \to 0\ ,
\end{aligned}$$

which implies that for any $\epsilon > 0$, there exists $N_0(\epsilon)\in\mathbb{N}$ such that $N > N_0(\epsilon)$ large enough

$$\mathbb{P}\left((E_{N,\eta_1})^c\right) < \frac{\epsilon}{2}\ . \tag{S4.13}$$

Note that λ_K^* is the unique maximizer of $G_1^*(\lambda)$, which implies

$$(G_1^*)'(\lambda_K^*) = \mathbb{E}[(Z\rho'((\lambda_K^*)^\top u_K(X)) - 1)u_K(X)] = 0 \ .$$

Then we have that for sufficiently large K:

$$\begin{aligned}
&\mathbb{E}[\|\hat{G}_1'(\lambda_K^*)\|^2] = \mathbb{E}[\hat{G}_1'(\lambda_K^*)^\top \hat{G}_1'(\lambda_K^*)] \\
=&\frac{1}{N^2}\sum_{i=1}^N \mathbb{E}\left[(Z_i\rho'((\lambda_K^*)^\top u_K(X_i)) - 1)^2 u_K(X_i)^\top u_K(X_i)\right] \\
&+ \frac{1}{N^2}\sum_{i\neq j} \mathbb{E}\left[(Z_i\rho'((\lambda_K^*)^\top u_K(X_i)) - 1)u_K(X_i)^\top \cdot (Z_j\rho'((\lambda_K^*)^\top u_K(X_j)) - 1)u_K(X_j)\right] \\
=&\frac{1}{N}\mathbb{E}[(Z\rho'((\lambda_K^*)^\top u_K(X)) - 1)^2 u_K(X)^\top u_K(X)] + \frac{N-1}{N}\times 0 \\
\leq& C_4^2\frac{1}{N}\mathbb{E}[u_K(X)^\top u_K(X)] = C_4^2\frac{K}{N} \ ,
\end{aligned} \tag{S4.14}$$

where $C_4^2 := \max\left\{1, \sup_{\gamma\in\Gamma_1}(\rho'(\gamma) - 1)^2\right\} < +\infty$.

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

$$\hat{\Lambda}_K(\epsilon) := \left\{\lambda \in \mathbb{R}^K : \|\lambda - \lambda_K^*\| \leq C_5(\epsilon)C_4\sqrt{\frac{K}{N}}\right\} \ . \tag{S4.15}$$

For $\forall\lambda \in \hat{\Lambda}_K(\epsilon),\ x \in \mathcal{X}$, and sufficiently large N, by Assumption 6, (S4.5) and (S4.8) we have

$$\begin{aligned}
&|\lambda^\top u_K(x) - (\lambda_K^*)^\top u_K(x)| \le \|\lambda - \lambda_K^*\|\|u_K(x)\| \le C_5(\epsilon) C_4 \sqrt{\frac{K}{N}}\zeta(K) \\
\Rightarrow\ & \lambda^\top u_K(x) \in \left[(\lambda_K^*)^\top u_K(x) - C_5(\epsilon)C_4\zeta(K)\sqrt{\frac{K}{N}},\ (\lambda_K^*)^\top u_K(x) + C_5(\epsilon)C_4\zeta(K)\sqrt{\frac{K}{N}}\right] \\
&\subset \Bigg[(\rho')^{-1}(\eta_1) - C_1K^{-\alpha} - C_2K^{-\alpha}\zeta(K) - C_5(\epsilon)C_4\zeta(K)\sqrt{\frac{K}{N}}, \\
&\qquad (\rho')^{-1}(\eta_2) + C_1K^{-\alpha} + C_2K^{-\alpha}\zeta(K) + C_5(\epsilon)C_4\zeta(K)\sqrt{\frac{K}{N}}\Bigg] \subset \Gamma_2(\epsilon)\ ,
\end{aligned} \tag{S4.16}$$

where

$$\Gamma_2(\epsilon) := [\underline{\gamma} - 1 - C_5(\epsilon), \overline{\gamma} + 1 + C_5(\epsilon)]\ ,$$

is a compact set independent of x.

By Mean Value Theorem, for any $\lambda \in \partial\hat{\Lambda}_K(\epsilon)$, there exists $\bar{\lambda}$ on the line joining λ and λ_K^* such that

$$\hat{G}_1(\lambda) = \hat{G}_1(\lambda_K^*) + (\lambda - \lambda_K^*)^\top \hat{G}_1'(\lambda_K^*) + \frac{1}{2}(\lambda - \lambda_K^*)^\top \hat{G}_1''(\bar{\lambda})(\lambda - \lambda_K^*)\ .$$

For the second order term, when N large enough we have

$$\begin{aligned}
(\lambda - \lambda_K^*)^\top \hat{G}_1''(\bar{\lambda})(\lambda - \lambda_K^*) &= \frac{1}{N}\sum_{i=1}^N Z_i \rho''(\bar{\lambda}^\top u_K(X_i))(\lambda - \lambda_K^*)^\top u_K(X_i) u_K(X_i)^\top (\lambda - \lambda_K^*) \\
&= \frac{1}{N}\sum_{i=1}^N Z_i \rho''(\bar{\lambda}^\top u_K(X_i))(\lambda - \lambda_K^*)^\top u_K(X_i) u_K(X_i)^\top (\lambda - \lambda_K^*) \\
&\le -\bar{b}(\epsilon)\frac{1}{N}\sum_{i=1}^N Z_i (\lambda - \lambda_K^*)^\top u_K(X_i) u_K(X_i)^\top (\lambda - \lambda_K^*)
\end{aligned}$$

$$= -\bar{b}(\epsilon) \cdot (\lambda - \lambda_K^*)^\top \hat{S}_N (\lambda - \lambda_K^*) \ , \tag{S4.17}$$

where $-\bar{b}(\epsilon) := \sup_{\gamma \in \Gamma_2(\epsilon)} \rho''(\gamma) < \infty$ because $\Gamma_2(\epsilon)$ is compact and ρ'' is a continuous function. Then on the event E_{N,η_1} with large enough N, we have that for any $\lambda \in \partial\hat{\Lambda}_K(\epsilon)$,

$$\begin{aligned}
&\hat{G}_1(\lambda) - \hat{G}_1(\lambda_K^*) \\
&= (\lambda - \lambda_K^*)^\top \hat{G}_1'(\lambda_K^*) + \frac{1}{2}(\lambda - \lambda_K^*)^\top \hat{G}_1''(\bar{\lambda})(\lambda - \lambda_K^*) \\
&\le \|\lambda - \lambda_K^*\| \|\hat{G}_1'(\lambda_K^*)\| - \frac{\bar{b}(\epsilon)}{2}(\lambda - \lambda_K^*)^\top \hat{S}_N (\lambda - \lambda_K^*) \\
&\le \|\lambda - \lambda_K^*\| \|\hat{G}_1'(\lambda_K^*)\| - \frac{\bar{b}(\epsilon)}{2}(\lambda - \lambda_K^*)^T \left(\mathbb{E}[f(1|X) u_K(X) u_K^\top(X)] - \frac{1}{2\eta_1} I_K \right) (\lambda - \lambda_K^*) \\
&\le \|\lambda - \lambda_K^*\| \|\hat{G}_1'(\lambda_K^*)\| - \frac{\bar{b}(\epsilon)}{2}(\lambda - \lambda_K^*)^\top \left(\frac{1}{\eta_1} I_K - \frac{1}{2\eta_1} I_K \right) (\lambda - \lambda_K^*) \\
&< \|\lambda - \lambda_K^*\| \left(\|\hat{G}_1'(\lambda_K^*)\| - \frac{\bar{b}(\epsilon)}{4\eta_1} \|\lambda - \lambda_K^*\| \right) \ ,
\end{aligned} \tag{S4.18}$$

where the first inequality follows from (S4.17). By Chebyshev's inequality and Inequality (S4.14), for sufficiently large N,

$$\mathbb{P}\left\{ \|\hat{G}_1'(\lambda_K^*)\| \ge \frac{\bar{b}(\epsilon)}{4\eta_1} \|\lambda - \lambda_K^*\| \right\} \le \frac{16\eta_1^2}{\bar{b}(\epsilon)^2 C_5^2(\epsilon)} \le \frac{\epsilon}{2} \ , \tag{S4.19}$$

the last inequality holds by choosing $C_5(\epsilon) \ge \dfrac{4\eta_1}{b(\epsilon)}\sqrt{\dfrac{2}{\epsilon}}$. Therefore, for sufficiently large N, by (S4.13) and (S4.19) we have that

$$\mathbb{P}\left((E_{N,\eta_1})^c \text{ or } \|\hat{G}_1'(\lambda_K^*)\| \ge \frac{\bar{b}(\epsilon)}{4\eta_1} \|\lambda - \lambda_K^*\| \right) \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

$$\Rightarrow \mathbb{P}\left(E_{N,\eta_1} \text{ and } \|\hat{G}_1'(\lambda_K^*)\| < \frac{\bar{b}(\epsilon)}{4\eta_1}\|\lambda - \lambda_K^*\| \right) > 1 - \epsilon \ . \tag{S4.20}$$

Then by (S4.18) and (S4.20), we get

$$\mathbb{P}\{\hat{G}_1(\lambda) - \hat{G}_1(\lambda_K^*) < 0, \ \forall \lambda \in \partial \hat{\Lambda}_K\} \geq 1 - \epsilon \ ,$$

for sufficiently large N. Note that the event set $\{\hat{G}_1(\lambda_K^*) > \hat{G}_1(\lambda), \ \forall \lambda \in \partial \hat{\Lambda}_K(\epsilon)\}$ implies that there exists a local maximum point in the interior of $\hat{\Lambda}_K(\epsilon)$. On the other hand, $\hat{G}_1$ is strictly concave function and $\hat{\lambda}_K$ is the unique global maximum point of $\hat{G}_1$, then we get

$$\mathbb{P}\left(\hat{\lambda}_K \in \hat{\Lambda}_K(\epsilon)\right) > 1 - \epsilon \ , \tag{S4.21}$$

i.e.

$$\|\hat{\lambda}_K - \lambda_K^*\| = O_p\left(\sqrt{\frac{K}{N}}\right) \ .$$

We next show that $\sup_{x\in\mathcal{X}} |\hat{w}_K(1|x) - w_K^*(1|x)| = O_p\left(\zeta(K)\sqrt{K/N}\right)$. By Mean Value Theorem, we can have

$$\begin{aligned} \hat{w}_K(1|x) - w_K^*(1|x) &= \rho'\left(\hat{\lambda}_K^\top u_K(x)\right) - \rho'\left((\lambda_K^*)^\top u_K(x)\right) \\ &= \rho''\left(\tilde{\lambda}_K^\top u_K(x)\right)(\hat{\lambda}_K - \lambda_K^*)^\top u_K(x) \ , \end{aligned}$$

where $\tilde{\lambda}_K$ lies on the line joining $\hat{\lambda}_K$ and λ_K^*. From (S4.21) and (S4.16), we have

$$\sup_{x\in\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_K^\top u_K(x)\right)\right| = O_p(1)\ , \tag{S4.22}$$

therefore, we can obtain that

$$\sup_{x\in\mathcal{X}}|\hat{w}_K(1|x)-w_K^*(1|x)| \leq \sup_{x\in\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_K^\top u_K(x)\right)\right|\cdot\|\hat{\lambda}_K-\lambda_K^*\|\cdot\sup_{x\in\mathcal{X}}\|u_K(x)\| = O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right)\ . \tag{S4.23}$$

Then we show $\int_{\mathcal{X}}|\hat{w}_K(1|x)-w_K^*(1|x)|^2dF_X(x) = O_p\left(K/N\right)$. By Mean Value Theorem, (S4.22), and the fact $\mathbb{E}\left[u_K(X)u_K(X)^\top\right]=I_K$, we have

$$\begin{aligned}
&\int_{\mathcal{X}}|\hat{w}_K(1|x)-w_K^*(1|x)|^2dF_X(x)\\
=&\int_{\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_K^\top u_K(x)\right)\cdot(\hat{\lambda}_K-\lambda_K^*)^\top u_K(x)\right|^2dF_X(x)\\
\leq&\sup_{x\in\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_K^\top u_K(x)\right)\right|^2\cdot(\hat{\lambda}_K-\lambda_K^*)^\top\cdot\int_{\mathcal{X}}u_K(x)u_K(x)^\top dF_X(x)\cdot(\hat{\lambda}_K-\lambda_K^*)\\
=&\sup_{x\in\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_K^\top u_K(x)\right)\right|^2\cdot\|\hat{\lambda}_K-\lambda_K^*\|^2\\
=&O_p(1)\cdot O_p\left(\frac{K}{N}\right)=O_p\left(\frac{K}{N}\right)\ .
\end{aligned}\tag{S4.24}$$

Similarly, we obtain that

$$\frac{1}{N}\sum_{i=1}^N|\hat{w}_K(1|X_i)-w_K^*(1|X_i)|^2$$

$$
\begin{aligned}
&= \frac{1}{N}\sum_{i=1}^{N}\left|\rho''\left(\tilde{\lambda}_K^\top u_K(X_i)\right)\cdot(\hat{\lambda}_K-\lambda_K^*)^\top u_K(X_i)\right| \\
&\leq \sup_{x\in\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_K^\top u_K(x)\right)\right|^2\cdot(\hat{\lambda}_K-\lambda_K^*)^\top\cdot\left\{\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)u_K(X_i)^\top\right\}(\hat{\lambda}_K-\lambda_K^*) \\
&\leq \sup_{x\in\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_K^\top u_K(x)\right)\right|^2\cdot\|\hat{\lambda}_K-\lambda_K^*\|^2\cdot\lambda_{\max}\left\{\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)u_K(X_i)^\top\right\} \\
&\leq O_p(1)\cdot O_p\left(\frac{K}{N}\right)\cdot O_p(1)=O_p\left(\frac{K}{N}\right)\ . \qquad \text{(S4.25)}
\end{aligned}
$$

Symmetrically, we have

$$
\begin{aligned}
&\|\hat{\beta}_K-\beta_K^*\| = O_p\left(\sqrt{\frac{K}{N}}\right)\ , \\
&\sup_{x\in\mathcal{X}}|\hat{w}_K(0|x)-w_K^*(0|x)| = O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right)\ , \\
&\int_{\mathcal{X}}|\hat{w}_K(0|x)-w_K^*(0|x)|^2 dF_X(x) = O_p\left(\frac{K}{N}\right)\ , \\
&\frac{1}{N}\sum_{i=1}^{N}|\hat{w}_K(0|X_i)-w_K^*(0|X_i)|^2 = O_p\left(\frac{K}{N}\right)\ .
\end{aligned}
$$

Next, we prove $\|\hat{\gamma}_K-\gamma_K^*\| = O_p\left(K^{-\alpha}+\sqrt{K/N}\right)$. This proof can be established by using a similar argument for showing $\|\hat{\lambda}_K-\lambda_K^*\| = O_p\left(\sqrt{K/N}\right)$. The whole proof works in parallel with minor modifications that we need to recompute the probability order of $\|\hat{H}'(\gamma_K^*)\|$ as what we have done in (S4.14). Note that

$$\|\hat{H}'(\gamma_K^*)\|$$

$$
\begin{aligned}
&= \left\| \frac{1}{N}\sum_{i=1}^{N} \rho_1'((\gamma_K^*)^\top u_K(X_i))u_K(X_i) - \frac{1}{N}\sum_{i=1}^{N} D_i \left\{Z_i \cdot \hat{w}_K(1|X_i) - (1-Z_i)\cdot \hat{w}_K(0|X_i)\right\} u_K(X_i) \right\| \\
&\le \left\| \frac{1}{N}\sum_{i=1}^{N} \rho_1'((\gamma_K^*)^\top u_K(X_i))u_K(X_i) - \frac{1}{N}\sum_{i=1}^{N} D_i \left\{ \frac{Z_i}{f(1|X_i)} - \frac{1-Z_i}{f(0|X_i)} \right\} u_K(X_i) \right\| \quad \text{(S4.26)} \\
&\quad + \left\| \frac{1}{N}\sum_{i=1}^{N} D_i Z_i \left\{ \frac{1}{f(1|X_i)} - \hat{w}_K(1|X_i) \right\} u_K(X_i) \right\| \quad \text{(S4.27)} \\
&\quad + \left\| \frac{1}{N}\sum_{i=1}^{N} D_i (1-Z_i) \left\{ \frac{1}{f(0|X_i)} - \hat{w}_K(0|X_i) \right\} u_K(X_i) \right\| . \quad \text{(S4.28)}
\end{aligned}
$$

For the term (S4.26), we have that for some finite constant $\tilde{C} > 0$,

$$
\begin{aligned}
\mathbb{E}\left[\|(\text{S4.26})\|^2\right] =& \frac{1}{N}\mathbb{E}\left[\left(\rho_1'((\gamma_K^*)^\top u_K(X)) - D\left\{ \frac{Z}{f(1|X)} - \frac{1-Z}{f(0|X)} \right\} \right)^2 \|u_K(X)\|^2 \right] \\
\le& \frac{\tilde{C}}{N} \cdot \mathbb{E}[\|u_K(X)\|^2] = \frac{\tilde{C}K}{N} , \quad \text{(the cross terms are zero since since } (H^*)'(\gamma_K^*) = 0)
\end{aligned}
$$

which implies (S4.26) $= O_p\left(\sqrt{K/N}\right)$. For the term (S4.27), by Cauchy-Schwartz' inequality we can deduce that

$$
\begin{aligned}
|(\text{S4.27})|^2 =& \left\{ \frac{1}{N}\sum_{i=1}^{N} D_i Z_i \left[\frac{1}{f(1|X_i)} - \hat{w}_K(1|X_i) \right] u_K(X_i)^\top \right\} \left\{ \frac{1}{N}\sum_{i=1}^{N} u_K(X_i)u_K(X_i)^\top \right\}^{-1} \\
&\cdot \left\{ \frac{1}{N}\sum_{j=1}^{N} u_K(X_j)u_K(X_j)^\top \right\}^2 \\
&\cdot \left\{ \frac{1}{N}\sum_{i=1}^{N} u_K(X_i)u_K(X_i) \right\}^{-1} \left\{ \frac{1}{N}\sum_{i=1}^{N} D_i Z_i \left[\frac{1}{f(1|X_i)} - \hat{w}_K(1|X_i) \right] u_K(X_i) \right\} \\
\le& \lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} u_K(X_i)u_K(X_i)^\top \right) \\
&\cdot \left\{ \frac{1}{N}\sum_{i=1}^{N} D_i Z_i \left[\frac{1}{f(1|X_i)} - \hat{w}_K(1|X_i) \right] u_K(X_i)^\top \right\} \left\{ \frac{1}{N}\sum_{i=1}^{N} u_K(X_i)u_K(X_i)^\top \right\}^{-1} \\
&\cdot \left\{ \frac{1}{N}\sum_{j=1}^{N} u_K(X_j)u_K(X_j)^\top \right\} \\
&\cdot \left\{ \frac{1}{N}\sum_{i=1}^{N} u_K(X_i)u_K(X_i) \right\}^{-1} \left\{ \frac{1}{N}\sum_{i=1}^{N} D_i Z_i \left[\frac{1}{f(1|X_i)} - \hat{w}_K(1|X_i) \right] u_K(X_i) \right\} \\
=& \lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} u_K(X_i)u_K(X_i)^\top \right) \cdot \frac{1}{N}\sum_{j=1}^{N} \hat{P}_N(X_j)\hat{P}_N(X_j)
\end{aligned}
$$

$$
\begin{aligned}
&\le \lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} u_K(X_i)u_K(X_i)^\top\right)\cdot\frac{1}{N}\sum_{j=1}^{N}\left\{D_iZ_i\left[\frac{1}{f(1|X_i)}-\hat{w}_K(1|X_i)\right]\right\}^2\\
&\le \lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} u_K(X_i)u_K(X_i)^\top\right)\cdot\frac{1}{N}\sum_{j=1}^{N}\left\{\frac{1}{f(1|X_i)}-\hat{w}_K(1|X_i)\right\}^2\\
&\le O_p(1)\cdot\left\{O(K^{-2\alpha})+O_p\left(\frac{K}{N}\right)\right\} \qquad \text{(by Lemmas 1 and 2)}\\
&\le O_p\left(K^{-2\alpha}+\frac{K}{N}\right)\ .
\end{aligned}
$$

where $\lambda_{\max}\left(N^{-1}\sum_{i=1}^{N} u_K(X_i)u_K(X_i)^\top\right)$ denotes the largest eigenvalue of the positive definite matrix $N^{-1}\sum_{i=1}^{N} u_K(X_i)u_K(X_i)^\top$, which is bounded with probability approaching to one; the second equality follows from the definition

$$
\hat{P}_N(X_j) := \left\{\frac{1}{N}\sum_{i=1}^{N} D_iZ_i\left[\frac{1}{f(1|X_i)}-\hat{w}_K(1|X_i)\right]u_K(X_i)^\top\right\}\left\{\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)u_K(X_i)^\top\right\}^{-1}u_K(X_j)\ ,
$$

and $\hat{P}_N(X_j)$ is the ordinary least squares (OLS) regression of $D_jZ_j\left[\frac{1}{f(1|X_j)}-\hat{p}(X_j)\right]$ on $u_K(X_j)$, then the second inequality follows from the geometric property of OLS:

$$
\frac{1}{N}\sum_{j=1}^{N}\hat{P}_N(X_j)\hat{P}_N(X_j)\le\frac{1}{N}\sum_{j=1}^{N}\left\{D_iZ_i\left[\frac{1}{f(1|X_i)}-\hat{w}_K(1|X_i)\right]\right\}^2\ .
$$

Therefore, we have that

$$
\text{(S4.27)} = O_p\left(K^{-\alpha}+\sqrt{\frac{K}{N}}\right)\ .
$$

Similarly, $|(\text{S4.28})| = O_p\left(K^{-\alpha} + \sqrt{\frac{K}{N}}\right)$. Therefore, we deduce

$$\|\hat{H}'(\gamma_K^*)\| \leq |(\text{S4.26})| + |(\text{S4.27})| + |(\text{S4.28})| = O_p\left(K^{-\alpha} + \sqrt{\frac{K}{N}}\right) .$$

Following the same argument as showing $\|\hat{\lambda}_K - \lambda_K^*\| = O_p\left(\sqrt{\frac{K}{N}}\right)$, we can obtain $\|\hat{\gamma}_K - \gamma_K^*\| = O_p\left(K^{-\alpha} + \sqrt{\frac{K}{N}}\right)$. Similar to obtain (S4.23), (S4.24), and (S4.25), we obtain

$$\begin{aligned}
&\sup_{x\in\mathcal{X}} |\hat{d}_K(x) - d_K^*(x)| = O_p\left(\zeta(K)K^{-\alpha} + \zeta(K)\sqrt{\frac{K}{N}}\right) , \\
&\int_{\mathcal{X}} |\hat{d}_K(x) - d_K^*(x)|^2 dF_X(x) = O_p\left(K^{-2\alpha} + \frac{K}{N}\right) , \\
&\frac{1}{N}\sum_{i=1}^{N} |\hat{d}_K(X_i) - d_K^*(X_i)|^2 = O_p\left(K^{-2\alpha} + \frac{K}{N}\right) .
\end{aligned}$$

□

Using Lemmas 1, 2 and triangular inequality, we get the proof of the proposition.

S5 Proof of Theorem 2

It suffices to show that the influence function of $\sqrt{N}(\hat{\tau} - \tau)$ is the same as the efficient influence function developed in Wang and Tchetgen Tchetgen (2018). We do some preliminary before the formal proof. Note that $\hat{\lambda}_K$

solves the following equation:

$$\frac{1}{N}\sum_{i=1}^{N} Z_i\rho'\left(\hat{\lambda}_K^\top u_K(X_i)\right)u_K(X_i) - \frac{1}{N}\sum_{i=1}^{N} u_K(X_i) = 0 \ .$$

By Mean Value Theorem, we have

$$\frac{1}{N}\sum_{i=1}^{N} Z_i\rho'\left((\lambda_K^*)^\top u_K(X_i)\right)u_K(X_i) - \frac{1}{N}\sum_{i=1}^{N} u_K(X_i) = -\frac{1}{N}\sum_{i=1}^{N} Z_i\rho''\left(\tilde{\lambda}_K^\top u_K(X_i)\right)u_K(X_i)u_K(X_i)^\top(\hat{\lambda}_K - \lambda_K^*)$$

where $\tilde{\lambda}_K$ lies on the line joining λ_K^* and $\hat{\lambda}_K$. Thus

$$\begin{aligned}\hat{\lambda}_K - \lambda_K^* = &-\left[\frac{1}{N}\sum_{i=1}^{N} Z_i\rho''\left((\lambda_K^*)^\top u_K(X_i)\right)u_K(X_i)u_K(X_i)^\top\right]^{-1} \\ &\cdot\left\{\frac{1}{N}\sum_{i=1}^{N} Z_i\rho'\left((\lambda_K^*)^\top u_K(X_i)\right)u_K(X_i) - \frac{1}{N}\sum_{i=1}^{N} u_K(X_i)\right\} \ .\end{aligned} \tag{S5.29}$$

Define $p_1^Y(X) := \mathbb{E}[Y|Z=1,X]$, $p_0^Y(X) := \mathbb{E}[Y|Z=0,X]$, and $\delta^Y(X) := p_1^Y(X) - p_0^Y(X)$. We also introduce the following notations:

$$\begin{aligned}\tilde{Q}_K(X) &= \tilde{\Psi}_K^\top\tilde{\Sigma}_K^{-1}u_K(X) \ , \\ \tilde{\Psi}_K &= -\int_{\mathcal{X}}\frac{p_1^Y(x)}{\delta^D(x)}f(1|x)\rho''(\tilde{\lambda}_K^\top u_K(x))u_K(x)dF_X(x) \ , \\ \tilde{\Sigma}_K &= \frac{1}{N}\sum_{i=1}^{N} Z_i\rho''(\tilde{\lambda}_K^\top u_K(X_i))u_K(X_i)u_K(X_i)^\top \ ,\end{aligned}$$

and

$$
\begin{aligned}
Q_K(X) &= \Psi_K^\top \Sigma_K^{-1} u_K(X)\ , \\
\Psi_K &= -\int_{\mathcal{X}} \frac{p_1^Y(x)}{\delta^D(x)} f(1|x) \rho''((\lambda_K^*)^\top u_K(x)) u_K(x) dF_X(x)\ , \\
\Sigma_K &= \mathbb{E}\left[f(1|X) \rho''((\lambda_K^*)^\top u_K(X)) u_K(X) u_K(X)^\top \right]\ .
\end{aligned}
$$

It should be noted that $Q_K(X)$ is the weighted L^2 projection of $-p_1^Y(X)/\delta^D(X)$ on the space spanned by $u_K(X)$.

Note that

$$
\sqrt{N}(\hat{\tau}-\tau) = \frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i \hat{w}_K(1|X_i) Y_i / \hat{d}_K(X_i) - \frac{1}{\sqrt{N}} \sum_{i=1}^N (1-Z_i) \hat{w}_K(0|X_i) Y_i / \hat{d}_K(X_i)\ .
$$

We shall derive the influence function of $\frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i \hat{w}_K(1|X_i) Y_i / \hat{d}_K(X_i)$, and similarly obtain that of $\frac{1}{\sqrt{N}} \sum_{i=1}^N (1-Z_i) \hat{w}_K(0|X_i) Y_i / \hat{d}_K(X_i)$. We can decompose $\frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i \hat{w}_K(1|X_i) Y_i / \hat{d}_K(X_i)$ as follows

$$
\begin{aligned}
&\frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i \hat{w}_K(1|X_i) Y_i / \hat{d}_K(X_i) \\
=&\frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{Z_i}{\hat{d}_K(X_i)} \{\hat{w}_K(1|X_i) - w_K^*(1|X_i)\} Y_i - \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{Z_i}{\delta^D(X_i)} \{\hat{w}_K(1|X_i) - w_K^*(1|X_i)\} Y_i \quad \text{(S5.30)} \\
&+ \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{Z_i}{\hat{d}_K(X_i)} w_K^*(1|X_i) Y_i - \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{Z_i}{\delta^D(X_i)} w_K^*(1|X_i) Y_i \quad \text{(S5.31)} \\
&+ \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{Z_i}{\delta^D(X_i)} \left(\hat{w}_K(1|X_i) - w_K^*(1|X_i)\right) Y_i - \int_{\mathcal{X}} \frac{p_1^Y(x) f(1|x)}{\delta^D(x)} (\hat{w}_K(1|x) - w_K^*(1|x)) dF_X(x) \right\} \\
&\quad \text{(S5.32)}
\end{aligned}
$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\left(w_K^*(1|X_i)-\frac{1}{f(1|X_i)}\right)\frac{Z_iY_i}{\delta^D(X_i)}-\mathbb{E}\left[\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\left(w_K^*(1|X)-\frac{1}{f(1|X)}\right)\right]\right\} \tag{S5.33}$$

$$+\sqrt{N}\mathbb{E}\left[\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\left(w_K^*(1|X)-\frac{1}{f(1|X)}\right)\right] \tag{S5.34}$$

$$+\sqrt{N}\int_{\mathcal{X}}\frac{p_1^Y(x)}{\delta^D(x)}f(1|x)(\hat{w}_K(1|x)-w_K^*(1|x))dF_X(x)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[Z_i\rho'((\lambda_K^*)^\top u_K(X_i))-1]\tilde{Q}_K(X_i) \tag{S5.35}$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[Z_i\rho'((\lambda_K^*)^\top u_K(X_i))-1](\tilde{Q}_K(X_i)-Q_K(X_i)) \tag{S5.36}$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{[Z_i\rho'((\lambda_K^*)^\top u_K(X_i))-1]Q_K(X_i)+\frac{p_1^Y(X_i)}{\delta^D(X_i)}\left(\frac{Z_i}{f(1|X_i)}-1\right)\right\} \tag{S5.37}$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{Z_iY_i}{f(1|X_i)\delta^D(X_i)}-\frac{p_1^Y(X_i)}{\delta^D(X_i)}\left(\frac{Z_i}{f(1|X_i)}-1\right)\right\}\ . \tag{S5.38}$$

We claim the following two lemmas, and their proofs are left in Sections S5.1 and S5.2.

Lemma 3. *Under Assumptions 1-6, the terms* (S5.30), (S5.32), (S5.33), (S5.34), (S5.35), (S5.36) *and* (S5.37) *are of* $o_p(1)$

Lemma 4. *Under Assumptions 1-6, the term* (S5.31) *has the following asymptotically equivalent linear expression*

$$\text{(S5.31)}=-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}D_i\cdot\frac{2Z_i-1}{f(Z_i|X_i)}\cdot\frac{p_1^Y(X_i)}{\delta^D(X_i)^2}+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{2Z_i-1}{\delta^D(X_i)^2}\cdot\frac{\mathbb{E}[D_i|Z_i,X_i]}{f(Z_i|X_i)}p_1^Y(X_i)+o_p(1)\ .$$

Using Lemmas 3 and 4, we have the following asymptotically equivalent linear expression for $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i\hat{w}_K(1|X_i)Y_i/\hat{d}_K(X_i)$:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i\hat{w}_K(1|X_i)Y_i/\hat{d}_K(X_i) \tag{S5.39}$$
$$=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{Z_iY_i}{f(1|X_i)\delta^D(X_i)}-\frac{p_1^Y(X_i)}{\delta^D(X_i)}\left(\frac{Z_i}{f(1|X_i)}-1\right)\right\}$$

$$-\frac{1}{\sqrt{N}}\sum_{i=1}^{N} D_i \cdot \frac{2Z_i-1}{f(Z_i|X_i)} \cdot \frac{p_1^Y(X_i)}{\delta^D(X_i)^2} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \frac{2Z_i-1}{\delta^D(X_i)^2} \cdot \frac{\mathbb{E}[D_i|Z_i,X_i]}{f(Z_i|X_i)} \cdot p_1^Y(X_i) + o_p(1)\ .$$

Symmetrically, we can also obtain

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-Z_i)\hat{w}_K(0|X_i)Y_i/\hat{d}_K(X_i) \qquad \text{(S5.40)}\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{(1-Z_i)Y_i}{f(0|X_i)\delta^D(X_i)} - \frac{p_0^Y(X_i)}{\delta^D(X_i)}\left(\frac{1-Z_i}{f(0|X_i)}-1\right)\right\}\\
&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N} D_i \cdot \frac{2Z_i-1}{f(Z_i|X_i)} \cdot \frac{p_0^Y(X_i)}{\delta^D(X_i)^2} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \frac{2Z_i-1}{\delta^D(X_i)^2} \cdot \frac{\mathbb{E}[D_i|Z_i,X_i]}{f(Z_i|X_i)} \cdot p_0^Y(X_i) + o_p(1)\ .
\end{aligned}$$

By combining (S5.39)and (S5.40), we can deduce that

$$\begin{aligned}
&\sqrt{N}(\hat{\tau}-\tau) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{Z_i\frac{\hat{w}_K(1|X_i)}{\hat{d}_K(X_i)}Y_i - (1-Z_i)\frac{\hat{w}_K(1|X_i)}{\hat{d}_K(X_i)}Y_i - \tau\right\}\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{2Z_i-1}{\delta^D(X_i)f(Z_i|X_i)}Y_i - \tau\right]\\
&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbb{E}[Y_i|Z_i=1,X_i]}{\delta^D(X_i)}\left\{\frac{Z_i}{f(1|X_i)}-1\right\} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbb{E}[Y_i|Z_i=0,X_i]}{\delta^D(X_i)}\left\{\frac{1-Z_i}{f(0|X_i)}-1\right\}\\
&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\delta(X_i)\left\{\frac{2Z_i-1}{f(Z_i|X_i)}\frac{D_i}{\delta^D(X_i)} - \frac{2Z_i-1}{f(Z_i|X_i)}\frac{\mathbb{E}[D_i|Z_i,X_i]}{\delta^D(X_i)}\right\} + o_p(1) \quad \left[\text{since } \delta(X) = \frac{\delta^Y(X)}{\delta^D(X)}\right]\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{2Z_i-1}{\delta^D(X_i)f(Z_i|X_i)}\left[Y_i - D_i\delta(X_i)\right] + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\delta(X_i) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tau\\
&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{Z_i\cdot\mathbb{E}[Y_i|Z_i=1,X_i]}{\delta^D(X_i)f(Z_i|X_i)} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{(1-Z_i)\cdot\mathbb{E}[Y_i|Z_i=0,X_i]}{\delta^D(X_i)f(Z_i|X_i)}\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{Z_i\cdot\mathbb{E}[D_i|Z_i=1,X_i]}{\delta^D(X_i)f(Z_i|X_i)}\cdot\delta(X_i) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{(1-Z_i)\cdot\mathbb{E}[D_i|Z_i=0,X_i]}{\delta^D(X_i)f(Z_i|X_i)}\cdot\delta(X_i) + o_p(1)\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{2Z_i-1}{\delta^D(X_i)f(Z_i|X_i)}\left[Y_i - D_i\delta(X_i)\right] + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\delta(X_i) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tau\\
&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{(2Z_i-1)\cdot\mathbb{E}[Y_i|Z_i=0,X_i]}{\delta^D(X_i)f(Z_i|X_i)} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{Z_i\cdot\mathbb{E}[Y_i|Z_i=0,X_i]}{\delta^D(X_i)f(Z_i|X_i)}\\
&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{Z_i\cdot\mathbb{E}[Y_i|Z_i=1,X_i]}{\delta^D(X_i)f(Z_i|X_i)} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{(2Z_i-1)\cdot\mathbb{E}[D_i|Z_i,X_i]}{\delta^D(X_i)f(Z_i|X_i)}\delta(X_i) + o_p(1)\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{2Z_i-1}{\delta^D(X_i)f(Z_i|X_i)}\left\{Y_i - D_i\delta(X_i) - \mathbb{E}[Y_i|Z_i=0,X_i]\right\} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\delta(X_i) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tau
\end{aligned}$$

$$
\begin{aligned}
&- \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{Z_i}{\delta^D(X_i) f(Z_i|X_i)} \delta^Y(X_i) + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{(2Z_i-1)\cdot \mathbb{E}[D_i|Z_i,X_i]}{\delta^D(X_i) f(Z_i|X_i)} \delta(X_i) + o_p(1) \\
=& \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{2Z_i-1}{\delta^D(X_i) f(Z_i|X_i)} \Big\{ Y_i - D_i\delta(X_i) - \mathbb{E}[Y_i|Z_i=0,X_i] \Big\} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \delta(X_i) - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau \\
&+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{2Z_i-1}{\delta^D(X_i) f(Z_i|X_i)} \delta(X_i) \left\{ \mathbb{E}[D_i|Z_i,X_i] - \frac{Z_i}{2Z_i-1} \delta^D(X_i) \right\} + o_p(1) \quad \left[\text{note } \delta(X) = \frac{\delta^Y(X)}{\delta^D(X)} \right] \\
=& \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{2Z_i-1}{\delta^D(X_i) f(Z_i|X_i)} \Big\{ Y_i - D_i\delta(X_i) - \mathbb{E}[Y_i|Z_i=0,X_i] \Big\} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \delta(X_i) - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau \\
&+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{2Z_i-1}{\delta^D(X_i) f(Z_i|X_i)} \delta(X_i) \Big\{ Z_i \cdot \mathbb{E}[D_i|Z_i=1,X_i] + (1-Z_i)\cdot \mathbb{E}[D_i|Z_i=0,X_i] \\
&\qquad - Z_i \delta^D(X_i) \Big\} + o_p(1) \qquad \left[\text{since } \frac{Z}{(2Z-1)} = Z \right] \\
=& \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{2Z_i-1}{\delta^D(X_i) f(Z_i|X_i)} \Big\{ Y_i - D_i\delta(X_i) - \mathbb{E}[Y_i|Z_i=0,X_i] \Big\} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \delta(X_i) - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau \\
&+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{2Z_i-1}{\delta^D(X_i) f(Z_i|X_i)} \delta(X_i) \cdot \mathbb{E}[D_i|Z_i=0,X_i] \\
=& \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \varphi_{eff}(D_i,Z_i,X_i,Y_i) + o_p(1)
\end{aligned}
$$

where

$$
\begin{aligned}
\varphi_{eff}(D_i,Z_i,X_i,Y_i) =& \frac{2Z_i-1}{f(Z_i|X_i)} \frac{1}{\delta^D(X_i)} \Big\{ Y_i - D_i\delta(X_i) - \mathbb{E}[Y_i|Z_i=0,X_i] \\
&+ \mathbb{E}[D_i|Z_i=0,X_i]\delta(X_i) \Big\} + \delta(X_i) - \tau \ ,
\end{aligned}
$$

is the efficient influence function developed in Wang and Tchetgen Tchetgen (2018).

S5.1 Proof of Lemma 3

For term (S5.30). By Proposition 1, Lemma 2, Assumptions 1 and 6, we deduce that

$$
\begin{aligned}
\text{(S5.30)} \leq & \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{|\delta^D(X_i) - \hat{d}_K(X_i)|}{|\hat{d}_K(X_i)\delta^D(X_i)|} \cdot |\hat{w}_K(1|X_i) - w_K^*(1|X_i)| \cdot |Y_i| \\
\leq & \sqrt{N} \cdot \sqrt{\frac{1}{N}\sum_{i=1}^{N} |\delta^D(X_i) - \hat{d}_K(X_i)|^2} \cdot \sqrt{\frac{1}{N}\sum_{i=1}^{N} |\hat{w}_K(1|X_i) - w_K^*(1|X_i)|^2 \cdot \frac{|Y_i|^2}{|\hat{d}_K(X_i)\delta^D(X_i)|^2}} \\
\leq & \sqrt{N} \cdot O_p(1) \cdot \sqrt{\frac{1}{N}\sum_{i=1}^{N} |\delta^D(X_i) - \hat{d}_K(X_i)|^2} \sqrt{\frac{1}{N}\sum_{i=1}^{N} \cdot \frac{|Y_i|^2}{|\hat{d}_K(X_i)\delta^D(X_i)|^2}} \cdot \sup_{x\in\mathcal{X}} |\hat{w}_K(1|x) - w_K^*(1|x)| \\
\leq & O_p(\sqrt{N}) \cdot O_p\left(K^{-\alpha} + \sqrt{\frac{K}{N}}\right) \cdot O_p(1) \cdot O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right) \\
= & O_p\left(\zeta(K)K^{-\alpha+\frac{1}{2}}\right) + O_p\left(\zeta(K)\sqrt{\frac{K^2}{N}}\right) \ .
\end{aligned}
$$

For term (S5.32). Using Mean Value Theorem twice, we can decompose (S5.32) as follows

$$
\begin{aligned}
\text{(S5.32)} = & \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\frac{Z_i}{\delta^D(X_i)} Y_i \rho''(\tilde{\lambda}_K^\top u_K(X_i)) u_K(X_i)^\top - \int_{\mathcal{X}} \frac{p_1^Y(x)}{\delta^D(x)} f(1|x) \rho''(\tilde{\lambda}_K^\top u_K(x)) u_K(x)^\top dF_X(x)\right](\hat{\lambda}_K - \lambda_K^*) \\
= & \left\{\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\frac{Z_i}{\delta^D(X_i)} Y_i \rho''((\lambda_K^*)^\top u_K(X_i)) u_K(X_i)^\top - \int_{\mathcal{X}} \frac{p_1^Y(x)}{\delta^D(x)} f(1|x) \rho''((\lambda_K^*)^\top u_K(x)) u_K(x)^\top dF_X(x)\right]\right. \\
& + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{Z_i}{\delta^D(X_i)} Y_i \rho'''(\xi_3(X_i))(\tilde{\lambda}_K - \lambda_K^*)^\top u_K(X_i) u_K(X_i)^\top \\
& \left. - \sqrt{N} \int_{\mathcal{X}} \frac{p_1^Y(x)}{\delta^D(x)} f(1|x) \rho'''(\xi_3(x))(\tilde{\lambda}_K - \lambda_K^*)^\top u_K(x) u_K(x)^\top dF_X(x)\right\}(\hat{\lambda}_K - \lambda_K^*) \\
= & (W_{1K} + W_{2K} + W_{3K})^\top (\hat{\lambda}_K - \lambda_K^*) \ ,
\end{aligned}
$$

where

$$
W_{1K} := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\frac{Z_i}{\delta^D(X_i)} Y_i \rho''((\lambda_K^*)^\top u_K(X_i)) u_K(X_i) - \int_{\mathcal{X}} \frac{p_1^Y(x)}{\delta^D(x)} f(1|x) \rho''((\lambda_K^*)^\top u_K(x)) u_K(x) dF_X(x)\right]
$$

$$W_{2K} := \frac{1}{\sqrt{N}}\left[\sum_{i=1}^{N}\frac{Z_i}{\delta^D(X_i)}Y_i\rho'''(\xi_3(X_i))u_K(X_i)u_K(X_i)^\top\right](\tilde{\lambda}_K-\lambda_K^*)$$

$$W_{3K} := -\sqrt{N}\int_{\mathcal{X}}\frac{p_1^Y(x)}{\delta^D(X_i)}f(1|x)\rho'''(\xi_3(x))u_K(x)u_K(x)^\top dF_X(x)(\tilde{\lambda}_K-\lambda_K^*)\ ,$$

where $\tilde{\lambda}_K$ lies on the line joining λ_K^* and $\hat{\lambda}_K$, $\xi_3(x)$ lies between $\tilde{\lambda}_K^\top u_K(x)$ and $(\lambda_K^*)^\top u_K(x)$.

Consider W_{1K}, for large enough K we have

$$\begin{aligned}
&\mathbb{E}[\|W_{1K}\|^2] = \mathbb{E}\left[\operatorname{tr}\left(W_{1K}^\top W_{1K}\right)\right] = \mathbb{E}\left[\frac{Z^2}{\delta^D(X)^2}Y^2\rho''((\lambda_K^*)^\top u_K(X))^2 u_K(X)^\top u_K(X)\right]\\
&\quad-\mathbb{E}\left[\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\rho''((\lambda_K^*)^\top u_K(X))u_K^\top(X)\right]\cdot\mathbb{E}\left[\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\rho''((\lambda_K^*)^\top u_K(X))u_K(X)\right]\\
&\le\mathbb{E}\left[\frac{1}{\delta^D(X)^2}\mathbb{E}[Y^2|X]\rho''((\lambda_K^*)^\top u_K(X))^2u_K(X)^\top u_K(X)\right]\\
&\le\sup_{\gamma\in\Gamma_1}|\rho''(\gamma)|^2\cdot\mathbb{E}\left[\frac{\mathbb{E}[Y^2|X]}{\delta^D(X)^2}\|u_K(X)\|^2\right]\le\sup_{\gamma\in\Gamma_1}|\rho''(\gamma)|^2\cdot\mathbb{E}\left[\frac{Y^2}{\delta^D(X)^2}\right]\zeta(K)^2\le O(\zeta(K)^2)\ .
\end{aligned}$$

where the last inequality holds because of the fact $\mathbb{E}\left[\frac{Y^2}{\delta^D(X)^2}\right]=\mathbb{E}\left[\frac{\delta^D(X)^2Y^2}{\delta^D(X)^4}\right]\le\eta_3^2\mathbb{E}\left[\frac{Y^2}{\delta^D(X)^4}\right]<\infty$ by Assumption 1. Therefore,

$$\|W_{1K}\| = O_p(\zeta(K))\ . \tag{S5.41}$$

Next, we calculate the stochastic order of W_{2K}.

$$\begin{aligned}
\|W_{2K}\| \le&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left|\frac{Y_i}{\delta^D(X_i)}\right|\cdot|\rho'''(\xi_3(X_i))|\cdot\|u_K(X_i)\|^2\cdot\|\tilde{\lambda}_K-\lambda_K^*\|\\
\le&\sup_{x\in\mathcal{X}}|\rho'''(\xi_3(x))|\sqrt{N}\cdot\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|\frac{Y_i}{\delta^D(X_i)}\right|^2}\cdot\sqrt{\frac{1}{N}\sum_{i=1}^{N}\|u_K(X_i)\|^4}\cdot\|\tilde{\lambda}_K-\lambda_K^*\|
\end{aligned}$$

$$\leq O_p(1) \cdot \sqrt{N} \cdot O_p(1) \cdot \sup_{x \in \mathcal{X}} \|u_K(x)\| \cdot \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|u_K(X_i)\|^2} \cdot O_p\left(\sqrt{K/N}\right)$$

$$\leq O_p(1) \cdot \sqrt{N} \cdot O_p(1) \cdot \zeta(K) \cdot O_p(\sqrt{K}) \cdot O_p\left(\sqrt{K/N}\right) = O_p\left(K\zeta(K)\right) \ , \tag{S5.42}$$

where the third inequality follows from the fact that $\sup_{x \in \mathcal{X}} |\rho'''(\xi_3(x))| = O_p(1)$, Assumption 1 and Lemma 2.

Finally, we consider W_{3K},

$$\begin{aligned}
&\left\| \int_{\mathcal{X}} f(1|x) \frac{p_1^Y(x)}{\delta^D(x)} \rho'''(\xi_3(x)) u_K(x) u_K(x)^\top dF(x) \right\| \\
\leq & \int_{\mathcal{X}} f(1|x) \left| \frac{p_1^Y(x)}{\delta^D(x)} \right| \cdot |\rho'''(\xi_3(x))| \cdot \|u_K(x)u_K(x)^\top\| dF_X(x) \\
\leq & \ O_p(1) \cdot \int_{\mathcal{X}} \left| \frac{p_1^Y(x)}{\delta^D(x)} \right| \|u_K(x)u_K(x)^\top\| dF_X(x) \\
\leq & \ O_p(1) \sqrt{\int_{\mathcal{X}} \left| \frac{p_1^Y(x)}{\delta^D(x)} \right|^2 dF_X(x)} \sqrt{\int_{\mathcal{X}} \|u_K(x)u_K(x)^\top\|^2 dF_X(x)} \\
= & O_p(\sqrt{K}\zeta(K)) \ ,
\end{aligned}$$

where the second inequality follows from the fact that $|\rho'''(\xi_3(x))| = O_p(1)$, and the last inequality follows from that

$$\int_{\mathcal{X}} \|u_K(x)u_K(x)^\top\|^2 dF_X(x) = \int_{\mathcal{X}} \mathrm{tr}(u_K(x)u_K(x)^\top u_K(x)u_K(x)^\top) dF_X(x) \leq \zeta(K)^2 K \ .$$

Note that $\tilde{\lambda}_K$ lies on the line joining $\hat{\lambda}_K$ and λ_K^*, by Lemma 2 $\|\tilde{\lambda}_K - \lambda_K^*\| =$

$O_p\left(\sqrt{\frac{K}{N}}\right)$, then we can deduce

$$\|W_{3K}\| = \sqrt{N} O_p\left(\sqrt{K}\zeta(K)\right) O_p\left(\sqrt{\frac{K}{N}}\right) = O_p(\zeta(K)K)\ . \tag{S5.43}$$

Therefore, by Lemma 2, (S5.41), (S5.42), (S5.43), (S5.42), and Assumption 6, we can obtain that

$$\begin{aligned}|(\text{S5.32})| \leq & (\|W_{1K}\| + \|W_{2K}\| + \|W_{3K}\|)\|\hat{\lambda}_K - \lambda_K^*\| \\ = & O_p(\zeta(K) + K\zeta(K) + K\zeta(K)) \cdot O_p\left(\sqrt{\frac{K}{N}}\right) = O_p\left(\zeta(K)\sqrt{\frac{K^3}{N}}\right)\ .\end{aligned}$$

For term **(S5.33).** The second moment of (S5.33) is

$$\begin{aligned}\mathbb{E}[|(\text{S5.33})|^2] = & \mathbb{E}\left[\left\{\left(w_K^*(1|X_i) - \frac{1}{f(1|X_i)}\right)\frac{Z_iY_i}{\delta^D(X_i)} - \mathbb{E}\left[\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\left(w_K^*(1|X) - \frac{1}{f(1|X)}\right)\right]\right\}^2\right] \\ \leq & \mathbb{E}\left[\left\{\left(w_K^*(1|X_i) - \frac{1}{f(1|X_i)}\right)\frac{Z_iY_i}{\delta^D(X_i)}\right\}^2\right] \\ \leq & \sup_{x\in\mathcal{X}}\left(w_K^*(1|x) - \frac{1}{f(1|x)}\right)^2 \mathbb{E}\left[\left\{\frac{Y_i}{\delta^D(X_i)}\right\}^2\right] \\ = & O(\zeta(K)^2 \cdot K^{-2\alpha}) \cdot O(1) = O(\zeta(K)^2 \cdot K^{-2\alpha})\ ,\end{aligned}$$

where the second equality follows from Lemma 1 and Assumption 1.

For term (S5.34)**.** By Lemma 1, Assumptions 1 and 6, we can derive

$$
\begin{aligned}
|(\text{S5.34})| =& \left|\sqrt{N}\mathbb{E}\left[\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\left(w_K^*(1|X)-\frac{1}{f(1|X)}\right)\right]\right| \\
\leq& \sqrt{N}\mathbb{E}\left[\left|\frac{p_1^Y(X)}{\delta^D(X)}\right|^2\right]^{\frac{1}{2}}\cdot\mathbb{E}\left[\left(w_K^*(1|X)-\frac{1}{f(1|X)}\right)^2\right]^{\frac{1}{2}} \\
\leq& \sqrt{N}\cdot O(1)\cdot O\left(K^{-\alpha}\right)=O(\sqrt{N}K^{-\alpha}) .
\end{aligned}
$$

For term (S5.35). By Mean Value Theorem and (S5.29), we can have

$$
\begin{aligned}
& \sqrt{N}\int_{\mathcal{X}}\frac{p_1^Y(x)}{\delta^D(x)}f(1|x)(\hat{w}_K(1|x)-w_K^*(1|x))dF_X(x) \qquad \text{(S5.44)} \\
=& \sqrt{N}\int_{\mathcal{X}}\frac{p_1^Y(x)}{\delta^D(x)}f(1|x)\rho''(\tilde{\lambda}_K^\top u_K(x))u_K(x)^\top dF_X(x)(\hat{\lambda}_K-\lambda_K^*) \\
=& -\sqrt{N}\tilde{\Psi}_K^\top(\hat{\lambda}_K-\lambda_K^*) \\
=& \frac{1}{\sqrt{N}}\sum_{i=1}^N[Z_i\rho'((\lambda_K^*)^\top u_K(X_i))-1]\tilde{\Psi}_K^\top\tilde{\Sigma}_K^{-1}u_K(X_i) \\
=& \frac{1}{\sqrt{N}}\sum_{i=1}^N[Z_i\rho'((\lambda_K^*)^\top u_K(X_i))-1]\tilde{Q}_K(X_i) ,
\end{aligned}
$$

then (S5.35) is identically equal to zero.

For term (S5.36)**.** Note that (S5.36) can be rewritten as

$$
\frac{1}{\sqrt{N}}\sum_{i=1}^N[Z_i\rho'((\lambda_K^*)^\top u_K(X_i))-1](\tilde{Q}_K(X_i)-Q_K(X_i))
$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [Z_i \rho'((\lambda_K^*)^\top u_K(X_i)) - 1](\tilde{\Psi}_K^\top \tilde{\Sigma}_K^{-1} - \Psi_K^\top \Sigma_K^{-1}) u_K(X_i)$$

$$= (\tilde{\Psi}_K^\top \tilde{\Sigma}_K^{-1} - \Psi_K^\top \Sigma_K^{-1}) \sqrt{N} \hat{G}_1'(\lambda_K^*)$$

$$= (\tilde{\Psi}_K - \Psi_K)^\top \tilde{\Sigma}_K^{-1} \sqrt{N} \hat{G}_1'(\lambda_K^*) + \Psi_K^\top (\tilde{\Sigma}_K^{-1} - \Sigma_K^{-1}) \sqrt{N} \hat{G}_1'(\lambda_K^*) \ . \quad \text{(S5.45)}$$

Consider the first term, by (S4.14) and Chebyshev's inequality, we get

$$\|\hat{G}_1'(\lambda_K^*)\| = O_p\left(\sqrt{\frac{K}{N}}\right) \ . \quad \text{(S5.46)}$$

By Mean Value Theorem we deduce that

$$\tilde{\Psi}_K - \Psi_K = -\int_{\mathcal{X}} \frac{p_1^Y(x)}{\delta^D(x)} f(1|x) [\rho''(\tilde{\lambda}_K^\top u_K(x)) - \rho''((\lambda_K^*)^\top u_K(x))] u_K(x) dF_X(x)$$
$$= \frac{W_{3K}}{\sqrt{N}} \ .$$

Note that $\lambda_{\min}(\tilde{\Sigma}_K^{-1}) = \dfrac{1}{\lambda_{\max}(\tilde{\Sigma}_K)} < 0$ and $\lambda_{\max}(\tilde{\Sigma}_K)$ is bounded away from zero with probability approaching to 1, we have $|\lambda_{\min}(\tilde{\Sigma}_K^{-1})| = O_p(1)$. Therefore, together with (S5.43) and (S5.46), we have

$$|(\tilde{\Psi}_K - \Psi_K)^\top \tilde{\Sigma}_K^{-1} \sqrt{N} \hat{G}_1'(\lambda_K^*)| \quad \text{(S5.47)}$$
$$\leq \|(W_{3K})^\top \tilde{\Sigma}_K^{-1}\| \|\hat{G}_1'(\lambda_K^*)\|$$
$$= \sqrt{W_{3K}^\top \left(\tilde{\Sigma}_K^{-1}\right)^2 W_{3K}} \cdot \|\hat{G}_1'(\lambda_K^*)\|$$

$$\leq \sqrt{\lambda_{\min}^2(\tilde{\Sigma}_K^{-1})W_{3K}^\top I_{K\times K}W_{3K}} \cdot \|\hat{G}_1'(\lambda_K^*)\|$$
$$\leq O_p(1)O_p(\zeta(K)K)O_p\left(\sqrt{\frac{K}{N}}\right) = O_p\left(\zeta(K)\sqrt{\frac{K^3}{N}}\right) .$$

Similarly, for the second term in (S5.45), we can deduce that

$$\begin{aligned}
&|\Psi_K^\top(\tilde{\Sigma}_K^{-1} - \Sigma_K^{-1})\sqrt{N}\hat{G}_1'(\lambda_K^*)| \qquad \text{(S5.48)}\\
=&\sqrt{N}|\hat{G}_1'(\lambda_K^*)^\top\tilde{\Sigma}_K^{-1}(\Sigma_K - \tilde{\Sigma}_K)\Sigma_K^{-1}\Psi_K|\\
\leq&\sqrt{N}\|\hat{G}_1'(\lambda_K^*)\|\cdot\|\Psi_K\|\cdot\|\tilde{\Sigma}_K^{-1}(\Sigma_K - \tilde{\Sigma}_K)\Sigma_K^{-1}\|\\
=&\sqrt{N}\|\hat{G}_1'(\lambda_K^*)\|\cdot\|\Psi_K\|\cdot\mathrm{tr}\left(\tilde{\Sigma}_K^{-1}(\Sigma_K - \tilde{\Sigma}_K)\Sigma_K^{-1}\Sigma_K^{-1}(\Sigma_K - \tilde{\Sigma}_K)\tilde{\Sigma}_K^{-1}\right)^{\frac{1}{2}}\\
=&\sqrt{N}\|\hat{G}_1'(\lambda_K^*)\|\cdot\|\Psi_K\|\cdot\mathrm{tr}\left(\Sigma_K^{-1}\Sigma_K^{-1}(\Sigma_K - \tilde{\Sigma}_K)\tilde{\Sigma}_K^{-1}\tilde{\Sigma}_K^{-1}(\Sigma_K - \tilde{\Sigma}_K)\right)^{\frac{1}{2}}\\
\leq&\sqrt{N}\|\hat{G}_1'(\lambda_K^*)\|\cdot\|\Psi_K\|\cdot\left|\lambda_{\max}(\Sigma_K^{-1}\Sigma_K^{-1})\right|^{\frac{1}{2}}\cdot\mathrm{tr}\left((\Sigma_K - \tilde{\Sigma}_K)\tilde{\Sigma}_K^{-1}\tilde{\Sigma}_K^{-1}(\Sigma_K - \tilde{\Sigma}_K)\right)^{\frac{1}{2}}\\
=&\sqrt{N}\|\hat{G}_1'(\lambda_K^*)\|\cdot\|\Psi_K\|\cdot\left|\lambda_{\max}(\Sigma_K^{-1}\Sigma_K^{-1})\right|^{\frac{1}{2}}\cdot\mathrm{tr}\left(\tilde{\Sigma}_K^{-1}\tilde{\Sigma}_K^{-1}(\Sigma_K - \tilde{\Sigma}_K)(\Sigma_K - \tilde{\Sigma}_K)\right)^{\frac{1}{2}}\\
\leq&\sqrt{N}\|\hat{G}_1'(\lambda_K^*)\|\cdot\|\Psi_K\|\cdot\left|\lambda_{\max}(\Sigma_K^{-1}\Sigma_K^{-1})\right|^{\frac{1}{2}}\cdot\left|\lambda_{\max}(\tilde{\Sigma}_K^{-1}\tilde{\Sigma}_K^{-1})\right|^{\frac{1}{2}}\cdot\mathrm{tr}\left((\Sigma_K - \tilde{\Sigma}_K)(\Sigma_K - \tilde{\Sigma}_K)\right)^{\frac{1}{2}}\\
=&\sqrt{N}\|\hat{G}_1'(\lambda_K^*)\|\cdot\|\Psi_K\|\cdot\left|\lambda_{\min}(\Sigma_K^{-1})\right|\cdot\left|\lambda_{\min}(\tilde{\Sigma}_K^{-1})\right|\cdot\left\|\Sigma_K - \tilde{\Sigma}_K\right\|
\end{aligned}$$

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

Consider $\left\|\Sigma_K - \tilde{\Sigma}_K\right\|$. Using Mean Value Theorem and triangle inequality, we can obtain that

$$\begin{aligned}
&\left\|\Sigma_K - \tilde{\Sigma}_K\right\| \qquad \text{(S5.49)}\\
\leq&\left\|\mathbb{E}\left[f(1|X)\rho''((\lambda_K^*)^\top u_K(X))u_K(X)u_K(X)^\top\right] - \frac{1}{N}\sum_{i=1}^N Z_i\rho''((\lambda_K^*)^\top u_K(X_i))u_K(X_i)u_K(X_i)^\top\right\|
\end{aligned}$$

$$
\begin{aligned}
&+\left\|\frac{1}{N}\sum_{i=1}^{N} Z_i\rho'''(\xi_4(X_i))u_K(X_i)u_K(X_i)^\top\cdot(\tilde{\lambda}_K-\lambda_K^*)^\top u_K(X_i)\right\| \\
\leq& O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right)+\left\|\frac{1}{N}\sum_{i=1}^{N} Z_i\rho'''(\xi_4(X_i))u_K(X_i)u_K(X_i)^\top\cdot(\tilde{\lambda}_K-\lambda_K^*)^\top u_K(X_i)\right\| .
\end{aligned}
$$

For the last item in the above expression, we can deduce that

$$
\begin{aligned}
&\left\|\frac{1}{N}\sum_{i=1}^{N} Z_i\rho'''(\xi_4(X_i))u_K(X_i)u_K(X_i)^\top\cdot(\tilde{\lambda}_K-\lambda_K^*)^\top u_K(X_i)\right\|^2 \\
=&\left\|\frac{1}{N}\sum_{i=1}^{N}\left\{Z_i\rho'''(\xi_4(X_i))u_K(X_i)(\tilde{\lambda}_K-\lambda_K^*)^\top u_K(X_i)\right\}u_K(X_i)^\top\right\|^2 \\
=&\mathrm{tr}\Bigg\{\frac{1}{N}\sum_{i=1}^{N}\left\{Z_i\rho'''(\xi_4(X_i))u_K(X_i)(\tilde{\lambda}_K-\lambda_K^*)^\top u_K(X_i)\right\}u_K(X_i)^\top\cdot\left[\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)u_K(X_i)^\top\right]^{-1} \\
&\quad\cdot\left[\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)u_K(X_i)^\top\right]^{2} \\
&\quad\cdot\left[\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)u_K(X_i)^\top\right]^{-1}\cdot\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)\left\{Z_i\rho'''(\xi_4(X_i))u_K(X_i)^\top(\tilde{\lambda}_K-\lambda_K^*)^\top u_K(X_i)\right\}\Bigg\} \\
\leq&\lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)u_K(X_i)^\top\right) \\
&\cdot\mathrm{tr}\Bigg\{\frac{1}{N}\sum_{i=1}^{N}\left\{Z_i\rho'''(\xi_4(X_i))u_K(X_i)(\tilde{\lambda}_K-\lambda_K^*)^\top u_K(X_i)\right\}u_K(X_i)^\top\cdot\left[\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)u_K(X_i)^\top\right]^{-1} \\
&\quad\cdot\left[\frac{1}{N}\sum_{j=1}^{N}u_K(X_j)u_K(X_j)^\top\right] \\
&\quad\cdot\left[\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)u_K(X_i)^\top\right]^{-1}\cdot\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)\left\{Z_i\rho'''(\xi_4(X_i))u_K(X_i)^\top(\tilde{\lambda}_K-\lambda_K^*)^\top u_K(X_i)\right\}\Bigg\} \\
=&\lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)u_K(X_i)^\top\right)\cdot\mathrm{tr}\left\{\frac{1}{N}\sum_{j=1}^{N}\tilde{L}_K(X_j)\tilde{L}_K(X_j)^\top\right\} \\
\leq&O_p(1)\cdot\mathrm{tr}\left\{\frac{1}{N}\sum_{j=1}^{N}\tilde{L}_K(X_j)\tilde{L}_K(X_j)^\top\right\} , \qquad \text{(S5.50)}
\end{aligned}
$$

where

$$
\tilde{L}_K(X_j)=\frac{1}{N}\sum_{i=1}^{N}\left\{Z_i\rho'''(\xi_4(X_i))u_K(X_i)(\tilde{\lambda}_K-\lambda_K^*)^\top u_K(X_i)\right\}u_K(X_i)^\top\cdot\left[\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)u_K(X_i)^\top\right]^{-1}u_K(X_j) ,
$$

is the OLS estimator of $\left\{Z_i\rho'''(\xi_4(X_j))u_K(X_j)(\tilde{\lambda}_K-\lambda_K^*)^\top u_K(X_j)\right\}$ on the basis $u_K(X_j)$, then it follows that

$$\begin{aligned}
&\mathrm{tr}\left\{\frac{1}{N}\sum_{j=1}^{N}\tilde{L}_K(X_j)\tilde{L}_K(X_j)^\top\right\} = \frac{1}{N}\sum_{j=1}^{N}\|\tilde{L}_K(X_j)\|^2 \leq \frac{1}{N}\sum_{j=1}^{N}\left\|Z_i\rho'''(\xi_4(X_j))u_K(X_j)(\tilde{\lambda}_K-\lambda_K^*)^\top u_K(X_j)\right\|^2 \\
&\leq \sup_{x\in\mathcal{X}}|\rho'''(\xi_4(x))|\cdot\zeta(K)^2\cdot(\tilde{\lambda}_K-\lambda_K^*)^\top\left\{\frac{1}{N}\sum_{j=1}^{N}u_K(X_j)u_K(X_j)^\top\right\}(\tilde{\lambda}_K-\lambda_K^*) \\
&\leq \sup_{x\in\mathcal{X}}|\rho'''(\xi_4(x))|\cdot\zeta(K)^2\cdot\|\tilde{\lambda}_K-\lambda_K^*\|^2\cdot\lambda_{\max}\left\{\frac{1}{N}\sum_{j=1}^{N}u_K(X_j)u_K(X_j)^\top\right\} \\
&\leq O_p(1)\cdot\zeta(K)^2\cdot O_p\left(\frac{K}{N}\right)\cdot O_p(1) = O_p\left(\zeta(K)^2\frac{K}{N}\right) \quad \text{(by Lemma 2)} . \qquad \text{(S5.51)}
\end{aligned}$$

Therefore, by combining (S5.49), (S5.50) and (S5.51), we obtain that

$$\left\|\Sigma_K - \tilde{\Sigma}_K\right\| = O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right) . \qquad \text{(S5.52)}$$

We next compute the order of $\Psi_K = -\mathbb{E}\left[\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\rho''((\lambda_K^*)^\top u_K(X))u_K(X)\right]$. By using the similar argument of obtaining (S4.6), we can deduce that

$$\begin{aligned}
\|\Psi_K\|^2 =&\mathbb{E}\left[\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\rho''((\lambda_K^*)^\top u_K(X))u_K(X)^\top\right]\mathbb{E}\left[\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\rho''((\lambda_K^*)^\top u_K(X))u_K(X)\right] \\
=&\mathbb{E}\left[\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\rho''((\lambda_K^*)^\top u_K(X))u_K(X)^\top\right]\cdot\mathbb{E}\left[u_K(X)u_K(X)^\top\right]^{-1} \\
&\cdot\mathbb{E}\left[u_K(X)u_K(X)^\top\right] \\
&\cdot\mathbb{E}\left[u_K(X)u_K(X)^\top\right]^{-1}\mathbb{E}\left[u_K(X)\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\rho''((\lambda_K^*)^\top u_K(X))\right] \\
=&\mathbb{E}\left[\left\{\mathbb{E}\left[\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\rho''((\lambda_K^*)^\top u_K(X))u_K(X)^\top\right]\cdot\mathbb{E}\left[u_K(X)u_K(X)^\top\right]^{-1}u_K(X)\right\}^2\right] \\
\leq&\mathbb{E}\left[\left\{\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\rho''((\lambda_K^*)^\top u_K(X))\right\}^2\right] \\
\leq& \sup_{x\in\mathcal{X}}|\rho''((\lambda_K^*)^\top u_K(x))|^2\cdot\left\|\frac{p_1^Y(X)}{\delta^D(X)}\right\|_{L^2}^2 \leq O(1) , \qquad \text{(S5.53)}
\end{aligned}$$

where the first inequality follows from the fact that

$$\mathbb{E}\left[\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\rho''((\lambda_K^*)^\top u_K(X))u_K(X)^\top\right]\cdot\mathbb{E}\left[u_K(X)u_K(X)^\top\right]^{-1}u_K(X)$$

is the L^2-projection of $\frac{p_1^Y(X)}{\delta^D(X)}f(1|X)\rho''((\lambda_K^*)^\top u_K(X))$ on the space spanned by $u_K(X)$. Combining (S5.48), (S5.52), and (S5.53) we can obtain

$$\begin{aligned}\|\Psi_K^\top(\tilde{\Sigma}_K^{-1}-\Sigma_K^{-1})\sqrt{N}\hat{G}_1'(\lambda_K^*)\| = &\ \sqrt{N}O_p\left(\sqrt{\frac{K}{N}}\right)O(1)O_p(1)O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right)\\ = &\ O_p\left(\zeta(K)\sqrt{\frac{K^2}{N}}\right),\end{aligned}$$

then together with (S5.47) and Assumption 6, we have

$$\begin{aligned}&\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N[Z_i\rho'((\lambda_K^*)^\top u_K(X_i))-1](\tilde{Q}_K(X_i)-Q_K(X_i))\right|\\ =&O_p\left(\zeta(K)\sqrt{\frac{K^3}{N}}\right)+O_p\left(\zeta(K)\sqrt{\frac{K^2}{N}}\right)=O_p\left(\zeta(K)\sqrt{\frac{K^3}{N}}\right).\end{aligned}$$

For term (S5.37)**:** Note that

$$\begin{aligned}&\mathbb{E}\left[\left([Z\rho'((\lambda_K^*)^\top u_K(X))-1]Q_K(X)+\frac{p_1^Y(X)}{f(1|X)\delta^D(X)}(Z-f(1|X))\right)^2\right]\\ \leq&\ 2\cdot\mathbb{E}\left[\left(\rho'((\lambda_K^*)^\top u_K(X))-\frac{1}{f(1|X)}\right)^2Z^2Q_K(X)^2\right]\end{aligned}$$

$$+ 2 \cdot \mathbb{E}\left[\left(\frac{p_1^Y(X)}{\delta^D(X)} + Q_K(X)\right)^2 \frac{1}{f(1|X)^2}(Z - f(1|X))^2\right] .$$

Consider the first term, by (S4.10) we have

$$\sup_{x\in\mathcal{X}}\left(\rho'((\lambda_K^*)^\top u_K(x)) - \frac{1}{f(1|x)}\right)^2 = O\left(\zeta(K)^2 K^{-2\alpha}\right) .$$

Note that $Q_K(X)$ can be written as follows:

$$Q_K(X) = A_K^\top u_K(X)$$

where A_K is defined by

$$\begin{aligned}
A_K :=& \arg\min_{A\in\mathbb{R}^K} \mathbb{E}\left[f(1|X)\left(-\rho''\left((\lambda_K^*)^\top u_K(X)\right)\right)\left\{\frac{p_1^Y(X)}{\delta^D(X)} - A^\top u_K(X)\right\}^2\right] \\
=& -\mathbb{E}\left[f(1|X)\rho''\left((\lambda_K^*)^\top u_K(X)\right)u_K(X)u_K(X)^\top\right]^{-1}\mathbb{E}\left[u_K(X)f(1|X)\rho''\left((\lambda_K^*)^\top u_K(X)\right)\frac{p_1^Y(X)}{\delta^D(X)}\right] .
\end{aligned}$$

Thus $Q_K(X)$ is the weighted L^2-projection of $-p_1^Y(X)/\delta^D(X)$ on the space spanned by $u_K(X)$. By definition and Assumption 4, we have that for any $A \in \mathbb{R}^K$,

$$\begin{aligned}
&\mathbb{E}\left[\left\{\frac{p_1^Y(X)}{\delta^D(X)} + Q_K(X)\right\}^2\right] \\
=&\mathbb{E}\left[\frac{f(1|X)\left(-\rho''\left((\lambda_K^*)^\top u_K(X)\right)\right)}{f(1|X)\left(-\rho''\left((\lambda_K^*)^\top u_K(X)\right)\right)}\left\{\frac{p_1^Y(X)}{\delta^D(X)} - A_K^\top u_K(X)\right\}^2\right] \\
\leq&\eta_1 \cdot \frac{1}{\inf_{\gamma\in\Gamma_1}|\rho''(\gamma)|} \cdot \mathbb{E}\left[f(1|X)\left(-\rho''\left((\lambda_K^*)^\top u_K(X)\right)\right)\left\{\frac{p_1^Y(X)}{\delta^D(X)} - A_K^\top u_K(X)\right\}^2\right] \\
\leq&\eta_1 \cdot \frac{1}{\inf_{\gamma\in\Gamma_1}|\rho''(\gamma)|} \cdot \mathbb{E}\left[f(1|X)\left(-\rho''\left((\lambda_K^*)^\top u_K(X)\right)\right)\left\{\frac{p_1^Y(X)}{\delta^D(X)} - A^\top u_K(X)\right\}^2\right]
\end{aligned}$$

$$\le \eta_1 \cdot \frac{\sup_{\gamma\in\Gamma_1}|\rho''(\gamma)|}{\inf_{\gamma\in\Gamma_1}|\rho''(\gamma)|} \cdot \mathbb{E}\left[\left\{\frac{p_1^Y(X)}{\delta^D(X)} - A^\top u_K(X)\right\}^2\right] .$$

Taking infimum over A in the above expression yields:

$$\mathbb{E}\left[\left\{\frac{p_1^Y(X)}{\delta^D(X)} + Q_K(X)\right\}^2\right] \le \eta_1 \cdot \frac{\sup_{\gamma\in\Gamma_1}|\rho''(\gamma)|}{\inf_{\gamma\in\Gamma_1}|\rho''(\gamma)|} \cdot \sup_{A\in\mathbb{R}^K} \mathbb{E}\left[\left\{\frac{p_1^Y(X)}{\delta^D(X)} - A^\top u_K(X)\right\}^2\right]$$
$$= O(K^{-2\alpha}) . \tag{S5.54}$$

Then by Lemma 1, Assumption 6 and above result, we obtain that

$$\begin{aligned}
&\mathbb{E}\left[\left([Z\rho'((\lambda_K^*)^\top u_K(X)) - 1]Q_K(X) + \frac{p_1^Y(X)}{\delta^D(X)f(1|X)}(Z - f(1|X))\right)^2\right] \\
\le& 2 \cdot \sup_{x\in\mathcal{X}}\left(\rho'((\lambda_K^*)^\top u_K(x)) - \frac{1}{f(1|x)}\right)^2 \cdot \mathbb{E}\left[Q_K(X)^2\right] + 2\eta_1^2 \cdot \mathbb{E}\left[\left\{\frac{p_1^Y(X)}{\delta^D(X)} + Q_K(X)\right\}^2\right] \\
\le& 2 \cdot \sup_{x\in\mathcal{X}}\left(\rho'((\lambda_K^*)^\top u_K(x)) - \frac{1}{f(1|x)}\right)^2 \cdot \mathbb{E}\left[Q_K(X)^2\right] + 2\eta_1^2 \cdot \mathbb{E}\left[\left\{\frac{p_1^Y(X)}{\delta^D(X)} + Q_K(X)\right\}^2\right] \\
=& O\left(K^{-2\alpha}\right) + O(K^{-2\alpha}) = O(K^{-2\alpha}) ,
\end{aligned}$$

which implies (S5.37) is of $O_p(K^{-\alpha})$ by Chebyshev's inequality.

Combining all all orders together and using Assumption 6, we can obtain that

$$\begin{aligned}
&(\text{S5.30}) + (\text{S5.32}) + (\text{S5.33}) + (\text{S5.34}) + (\text{S5.35}) + (\text{S5.36}) + (\text{S5.37}) \\
=& \left\{O_p\left(\zeta(K)K^{-\alpha+\frac{1}{2}}\right) + O_p\left(\zeta(K)\sqrt{\frac{K^2}{N}}\right)\right\} + O_p\left(\zeta(K)\sqrt{\frac{K^3}{N}}\right) + O(\zeta(K)^2 \cdot K^{-2\alpha}) \\
&+ O(\sqrt{N}K^{-\alpha}) + 0 + O_p\left(\zeta(K)\sqrt{\frac{K^3}{N}}\right) + O_p(K^{-\alpha}) = o_p(1) .
\end{aligned}$$

Therefore, we can obtain that

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i \hat{w}_K(1|X_i)Y_i/\hat{d}_K(X_i) \\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \frac{Z_i}{\hat{d}_K(X_i)} w_K^*(1|X_i)Y_i - \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \frac{Z_i}{\delta^D(X_i)} w_K^*(1|X_i)Y_i \\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \left\{ \frac{Z_iY_i}{f(1|X_i)\delta^D(X_i)} - \frac{p_1^Y(X_i)}{\delta^D(X_i)}\left(\frac{Z_i}{f(1|X_i)}-1\right)\right\} + o_p(1)\ .
\end{aligned}
$$

S5.2 Proof of Lemma 4

Note that

$$
\begin{aligned}
\text{(S5.31)} =&\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \frac{Z_i}{\hat{d}_K(X_i)} w_K^*(1|X_i)Y_i - \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \frac{Z_i}{\delta^D(X_i)} w_K^*(1|X_i)Y_i \\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i \frac{\delta^D(X_i)-\hat{d}_K(X_i)}{\hat{d}_K(X_i)\delta^D(X_i)}\left\{ w_K^*(1|X_i) - \frac{1}{f(1|X_i)}\right\} Y_i \\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i \frac{\delta^D(X_i)-\hat{d}_K(X_i)}{\hat{d}_K(X_i)\delta^D(X_i)} \frac{Y_i}{f(1|X_i)} \\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i \frac{\delta^D(X_i)-\hat{d}_K(X_i)}{\hat{d}_K(X_i)\delta^D(X_i)}\left\{ w_K^*(1|X_i) - \frac{1}{f(1|X_i)}\right\} Y_i \quad \text{(S5.55)} \\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i \left\{ \frac{\delta^D(X_i)-\hat{d}_K(X_i)}{\hat{d}_K(X_i)\delta^D(X_i)} - \frac{\delta^D(X_i)-\hat{d}_K(X_i)}{\delta^D(X_i)^2}\right\} \frac{Y_i}{f(1|X_i)} \\
&\hfill \text{(S5.56)} \\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i \left\{ \frac{\delta^D(X_i)-\hat{d}_K(X_i)}{\delta^D(X_i)^2}\right\} \frac{Y_i}{f(1|X_i)}\ . \quad \text{(S5.57)}
\end{aligned}
$$

We first prove that (S5.55) and (S5.56) are of $o_p(1)$.

For the term (S5.55). Using Cauchy-Schwartz's inequality, Lemmas 1 and 2, and Assumption 6, we can deduce that

$$
\begin{aligned}
|(\text{S5.55})| \leq & \frac{1}{\sqrt{N}} \sum_{i=1}^{N} |\delta^D(X_i) - \hat{d}_K(X_i)| \cdot \left| w_K^*(1|X_i) - \frac{1}{f(1|X_i)} \right| \cdot \left| \frac{Y_i}{\hat{d}_K(X_i)\delta^D(X_i)} \right| \\
\leq & \sqrt{N} \sup_{x \in \mathcal{X}} \left| w_K^*(1|x) - \frac{1}{f(1|x)} \right| \cdot \sqrt{\frac{1}{N} \sum_{i=1}^{N} |\delta^D(X_i) - \hat{d}_K(X_i)|^2} \cdot \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left| \frac{Y_i}{\hat{d}_K(X_i)\delta^D(X_i)} \right|^2} \\
\leq & \sqrt{N} \cdot O(\zeta(K)K^{-\alpha}) \cdot O_p\left(K^{-\alpha} + \sqrt{\frac{K}{N}} \right) \cdot O_p(1) \\
= & O_p(\sqrt{N}\zeta(K)K^{-2\alpha}) + O_p\left(\zeta(K)K^{-\alpha+\frac{1}{2}} \right) = o_p(1) \ .
\end{aligned}
$$

For the term (S5.56). Using Lemma 2, Cauchy-Schwartz's inequality, Assumptions 1, 4 and 6, we can deduce that

$$
\begin{aligned}
|(\text{S5.56})| \leq & \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{f(1|X_i)} \cdot |\delta^D(X_i) - \hat{d}_K(X_i)|^2 \cdot \frac{1}{|\hat{d}_K(X_i)\delta^D(X_i)|} \cdot \frac{|Y_i|}{|\delta^D(X_i)|} \\
\leq & \sqrt{N} \cdot \eta_1 \cdot \sup_{x \in \mathcal{X}} |\delta^D(x) - \hat{d}_K(x)|^2 \cdot \left[\frac{1}{N} \sum_{i=1}^{N} \frac{1}{|\hat{d}_K(X_i)|} \cdot \frac{|Y_i|}{|\delta^D(X_i)|^2} \right] \\
\leq & \sqrt{N} \cdot \eta_1 \cdot \sup_{x \in \mathcal{X}} |\delta^D(x) - \hat{d}_K(x)|^2 \cdot \left[\frac{1}{N} \sum_{i=1}^{N} \frac{1}{|\hat{d}_K(X_i)|^2} \right]^{\frac{1}{2}} \cdot \left[\frac{1}{N} \sum_{i=1}^{N} \frac{|Y_i|^2}{|\delta^D(X_i)|^4} \right]^{\frac{1}{2}} \\
\leq & O(\sqrt{N}) \cdot O_p\left(\zeta(K)^2 K^{-2\alpha} + \zeta(K)^2 \frac{K}{N} \right) \cdot O_p(1) \cdot O_p(1) \\
= & O_p\left(\sqrt{N}\zeta(K)^2 K^{-2\alpha} + \zeta(K)^2 \frac{K}{\sqrt{N}} \right) = o_p(1) \ .
\end{aligned}
$$

For the term (S5.57). Similar to the definition of $\tilde{Q}_K(X)$ and $Q_K(X)$ in P.19, we define the following notation

$$\tilde{R}_K(X) = \left[\int_{\mathcal{X}} \rho_1''(\tilde{\gamma}_K^\top u_K(x)) \frac{p_1^Y(x)}{\delta^D(x)^2} u_K^\top(x) dF_X(x)\right] \left[-\frac{1}{N}\sum_{i=1}^N \rho_1''\left(\tilde{\gamma}_K^\top u_K(X_i)\right) u_K(X_i) u_K(X_i)^\top\right]^{-1} u_K(X) ,$$
$$R_K(X) = \left[\int_{\mathcal{X}} \rho_1''((\gamma_K^*)^\top u_K(x)) \frac{p_1^Y(x)}{\delta^D(x)^2} u_K^\top(x) dF_X(x)\right] \left[-\int_{\mathcal{X}} \rho_1''\left((\gamma_K^*)^\top u_K(x)\right) u_K(x) u_K(x)^\top dF_X(x)\right]^{-1} u_K(X) ,$$

where $R_K(X)$ is the weighted L^2 projection of $-\frac{p_1^Y(X)}{\delta^D(X)^2}$ on the space spanned by $u_K(X)$, $\tilde{\gamma}_K$ lies on the line joining $\hat{\gamma}_K$ and γ_K^* such that the following Mean Value Theorem holds:

$$\frac{1}{N}\sum_{i=1}^N \rho_1'((\gamma_K^*)^\top u_K(X_i)) u_K(X_i) - \sum_{i=1}^N D_i \left\{Z_i \hat{w}_K(1|X_i) - (1-Z_i)\hat{w}_K(0|X_i)\right\} u_K(X_i)$$
$$= -\frac{1}{N}\sum_{i=1}^N \rho_1'(\tilde{\gamma}_K^\top u_K(X_i)) u_K(X_i) u_K(X_i)^\top (\hat{\gamma}_K - \gamma_K^*) \ .$$

We can decompose (S5.57) as follows:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N Z_i \left\{\hat{d}_K(X_i) - \delta^D(X_i)\right\} \frac{Y_i}{\delta^D(X_i)^2 f(1|X_i)}$$
$$= \frac{1}{\sqrt{N}}\sum_{i=1}^N \left[\frac{\{\rho_1'(\tilde{\gamma}_K^\top u_K(X_i)) - \rho_1'((\gamma_K^*)^\top u_K(X_i))\}}{\delta^D(X_i)^2 f(1|X_i)} Z_i Y_i - \int_{\mathcal{X}} \left\{\rho_1'(\hat{\gamma}_K^\top u_K(x)) - \rho_1'((\gamma_K^*)^\top u_K(x))\right\} \frac{p_1^Y(x)}{\delta^D(x)^2} dF_X(x)\right] \tag{S5.58}$$
$$+ \frac{1}{\sqrt{N}}\sum_{i=1}^N \left\{\left\{\rho_1'((\gamma_K^*)^\top u_K(X_i)) - \delta^D(X_i)\right\} \frac{Z_i Y_i}{\delta^D(X_i)^2 f(1|X_i)} - \mathbb{E}\left[\left\{\rho_1'((\gamma_K^*)^\top u_K(X)) - \delta^D(X)\right\} \frac{p_1^Y(X)}{\delta^D(X)^2}\right]\right\} \tag{S5.59}$$
$$+ \sqrt{N}\mathbb{E}\left[\left\{\rho_1'((\gamma_K^*)^\top u_K(X)) - \delta^D(X)\right\} \frac{p_1^Y(X)}{\delta^D(X)^2}\right] \tag{S5.60}$$
$$+ \frac{1}{\sqrt{N}}\left\{\int_{\mathcal{X}} \left\{\rho_1'(\hat{\gamma}_K^\top u_K(x)) - \rho_1'((\gamma_K^*)^\top u_K(x))\right\} \frac{p_1^Y(x)}{\delta^D(x)^2} dF_X(x)\right.$$

$$- \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\rho_1'((\gamma_K^*)^\top u_K(X_i)) - D_i\{Z_i \cdot \hat{w}_K(1|X_i) - (1-Z_i)\hat{w}_K(0|X_i)\} \right] \tilde{R}_K(X_i) \Bigg\} \tag{S5.61}$$

$$+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\rho_1'((\gamma_K^*)^\top u_K(X_i)) - D_i\{Z_i \cdot \hat{w}_K(1|X_i) - (1-Z_i)\hat{w}_K(0|X_i)\} \right] \cdot \left[\tilde{R}_K(X_i) - R_K(X_i) \right] \tag{S5.62}$$

$$+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[D_i \left\{ Z_i \cdot w_K^*(1|X_i) - (1-Z_i) q^*(X_i) \right\} - D_i\{Z_i \cdot \hat{w}_K(1|X_i) - (1-Z_i)\hat{w}_K(0|X_i)\} \right] \cdot R_K(X_i) \tag{S5.63}$$

$$+ \Bigg\{ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\rho_1'((\gamma_K^*)^\top u_K(X_i)) - D_i \left\{ Z_i \cdot w_K^*(1|X_i) - (1-Z_i) q^*(X_i) \right\} \right] \cdot R_K(X_i) + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\delta^D(X_i) - D_i \left\{ \frac{Z_i}{f(1|X_i)} - \frac{1-Z_i}{f(0|X_i)} \right\} \right] \frac{p_1^Y(X_i)}{\delta^D(X_i)^2} \Bigg\} \tag{S5.64}$$

$$- \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\delta^D(X_i) - D_i \left\{ \frac{Z_i}{f(1|X_i)} - \frac{1-Z_i}{f(0|X_i)} \right\} \right] \frac{p_1^Y(X_i)}{\delta^D(X_i)^2} . \tag{S5.65}$$

Using the similar argument as for establishing Lemma 3, we can show that (S5.58), (S5.59), (S5.60), (S5.61), (S5.62), and (S5.64) are of $o_p(1)$; indeed,

- "Proof of (S5.58) is of $o_p(1)$" is analogous to "Proof of (S5.32) is of $o_p(1)$";
- "Proof of (S5.59) is of $o_p(1)$" is analogous to "Proof of (S5.33) is of $o_p(1)$";
- "Proof of (S5.60) is of $o_p(1)$" is analogous to "Proof of (S5.34) is of $o_p(1)$";
- "Proof of (S5.61) is of $o_p(1)$" is analogous to "Proof of (S5.35) is of $o_p(1)$";
- "Proof of (S5.62) is of $o_p(1)$" is analogous to "Proof of (S5.36) is of $o_p(1)$";
- "Proof of (S5.64) is of $o_p(1)$" is analogous to "Proof of (S5.37) is of $o_p(1)$";

For the term (S5.63), we can have that

$$\text{(S5.63)} = -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[D_iZ_i\left\{\hat{w}_K(1|X_i) - w_K^*(1|X_i)\right\}\right]\cdot R_K(X_i) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[D_i(1-Z_i)\left\{\hat{w}_K(0|X_i) - w_K^*(0|X_i)\right\}\right]\cdot R_K(X_i)\ .$$

We make the following claims:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[D_iZ_i\left\{\hat{w}_K(1|X_i) - w_K^*(1|X_i)\right\}\right]\cdot R_K(X_i) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbb{E}[D|Z=1,X_i]}{\delta^D(X_i)^2}\left\{\frac{Z_i}{f(1|X_i)} - 1\right\}p_1^Y(X_i) + o_p(1)\ , \tag{S5.66}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[D_i(1-Z_i)\left\{\hat{w}_K(0|X_i) - w_K^*(0|X_i)\right\}\right]\cdot R_K(X_i) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbb{E}[D|Z=0,X_i]}{\delta^D(X_i)^2}\left\{\frac{1-Z_i}{f(0|X_i)} - 1\right\}p_1^Y(X_i) + o_p(1)\ . \tag{S5.67}$$

We prove the claim (S5.66), and by the similar argument we can obtain (S5.67).

For the claim (S5.66). Note that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[D_iZ_i\left\{\hat{w}_K(1|X_i) - w_K^*(1|X_i)\right\}\right]\cdot R_K(X_i)$$

$$= -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[D_iZ_i\left\{\hat{w}_K(1|X_i) - w_K^*(1|X_i)\right\}\right]\cdot\frac{p_1^Y(X_i)}{\delta^D(X_i)^2} \tag{S5.68}$$

$$+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[D_iZ_i\left\{\hat{w}_K(1|X_i) - w_K^*(1|X_i)\right\}\right]\cdot\left\{\frac{p_1^Y(X_i)}{\delta^D(X_i)^2} + R_K(X_i)\right\}\ . \tag{S5.69}$$

For the term (S5.69), by Lemma 2 and Cauchy-Schwartz's inequality, we deduce that

$$
\begin{aligned}
&\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[D_iZ_i\left\{\hat{w}_K(1|X_i)-w_K^*(1|X_i)\right\}\right]\cdot\left\{\frac{p_1^Y(X_i)}{\delta^D(X_i)^2}+R_K(X_i)\right\}\right|\\
\leq&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left|\hat{w}_K(1|X_i)-w_K^*(1|X_i)\right|\cdot\left|\frac{p_1^Y(X_i)}{\delta^D(X_i)^2}+R_K(X_i)\right|\\
\leq&\sqrt{N}\cdot\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|\hat{w}_K(1|X_i)-w_K^*(1|X_i)\right|^2}\cdot\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|\frac{p_1^Y(X_i)}{\delta^D(X_i)^2}+R_K(X_i)\right|^2}\\
\leq&\sqrt{N}\cdot O_p\left(\sqrt{\frac{K}{N}}\right)\cdot O_p(K^{-\alpha})=o_p(1)\ .
\end{aligned}
$$

where the third inequality follows from that fact that $R_K(X)$ is the weighted L^2-projection of $-p_1^Y(X)/\delta^D(X)^2$ on the space spanned by $u_K(X)$ (with the weights $-\rho_1''((\gamma_K^*)^\top u_K(X))$).

For the term (S5.68), we have the following decomposition:

$$
\begin{aligned}
\text{(S5.68)}=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Bigg[-D_iZ_i\left\{\hat{w}_K(1|X_i)-w_K^*(1|X_i)\right\}\cdot\frac{p_1^Y(X_i)}{\delta^D(X_i)^2}\\
&\qquad+\int_{\mathcal{X}}\frac{p_1^Y(x)}{\delta^D(x)^2}\cdot f(1|x)\cdot\mathbb{E}[D|Z=1,X=x]\left\{\hat{w}_K(1|x)-w_K^*(1|x)\right\}dF_X(x)\Bigg] \quad \text{(S5.70)}\\
&+\Bigg[-\sqrt{N}\int_{\mathcal{X}}\frac{p_1^Y(x)}{\delta^D(x)^2}\cdot f(1|x)\cdot\mathbb{E}[D|Z=1,X=x]\left\{\hat{w}_K(1|x)-w_K^*(1|x)\right\}dF_X(x)\\
&\qquad+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(Z_i\rho'\left((\lambda_K^*)^\top u_K(X_i)\right)-1\right)\tilde{P}_K(X_i)\Bigg] \quad \text{(S5.71)}\\
&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(Z_i\rho'\left((\lambda_K^*)^\top u_K(X_i)\right)-1\right)\left\{\tilde{P}_K(X_i)-P_K(X_i)\right\} \quad \text{(S5.72)}
\end{aligned}
$$

$$-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(Z_i\rho'\left((\lambda_K^*)^\top u_K(X_i)\right)-1\right)P_K(X_i)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbb{E}[D|Z=1,X_i]}{\delta^D(X_i)^2}\left\{\frac{Z_i}{f(1|X_i)}-1\right\}p_1^Y(X_i) \tag{S5.73}$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbb{E}[D|Z=1,X_i]}{\delta^D(X_i)^2}\left\{\frac{Z_i}{f(1|X_i)}-1\right\}p_1^Y(X_i)\ , \tag{S5.74}$$

where

$$\begin{aligned}\tilde{P}_K(X_i)=&\left[\int_{\mathcal{X}}\frac{p_1^Y(x)}{\delta^D(x)^2}\cdot f(1|x)\cdot\mathbb{E}[D|Z=1,X=x]\cdot\rho''(\tilde{\lambda}_K^\top u_K(x))u_K(x)^\top dF_X(x)\right]\\&\cdot\left[-\frac{1}{N}\sum_{i=1}^{N}Z_i\rho''\left(\tilde{\lambda}_K^\top u_K(X_i)\right)u_K(X_i)u_K(X_i)^\top\right]^{-1}\\&\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_K(X_i)\left\{Z_i\rho'\left((\lambda_K^*)^\top u_K(X_i)\right)-1\right\}\end{aligned}$$

and

$$\begin{aligned}P_K(X_i)=-&\left[\int_{\mathcal{X}}\frac{p_1^Y(x)}{\delta^D(x)^2}\cdot f(1|x)\cdot\mathbb{E}[D|Z=1,X=x]\cdot\rho''((\lambda_K^*)^\top u_K(x))u_K(x)^\top dF_X(x)\right]\\&\cdot\mathbb{E}\left[f(1|X)\rho''\left((\lambda_K^*)^\top u_K(X_i)\right)u_K(X_i)u_K(X_i)^\top\right]^{-1}\\&\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_K(X_i)\left\{Z_i\rho'\left((\lambda_K^*)^\top u_K(X_i)\right)-1\right\}\ .\end{aligned}$$

and $P_K(X)$ is the weighted L^2-projection of $-p_1^Y(X)/\delta^D(X)^2$ on the space spanned by $u_K(X)$ (with the weights $-\rho''((\lambda_K^*)^\top u_K(X)))$. Using the similar argument for establishing Lemma 3, we can prove that (S5.70), (S5.71), (S5.72), and (S5.73) are of $o_p(1)$. Indeed,

- "Proof of (S5.70) is of $o_p(1)$" is analogous to "Proof of (S5.32) is of $o_p(1)$";

- "Proof of (S5.71) is of $o_p(1)$" is analogous to "Proof of (S5.35) is of $o_p(1)$";

- "Proof of (S5.72) is of $o_p(1)$" is analogous to "Proof of (S5.36) is of $o_p(1)$";

- "Proof of (S5.73) is of $o_p(1)$" is analogous to "Proof of (S5.37) is of $o_p(1)$";

Therefore, we can obtain the claim (S5.66). Similarly, we can obtain the claim (S5.67). Then we have that

$$\begin{aligned}\text{(S5.63)} =& -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbb{E}[D|Z=1,X_i]}{\delta^D(X_i)^2}\left\{\frac{Z_i}{f(1|X_i)}-1\right\}p_1^Y(X_i)+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbb{E}[D|Z=0,X_i]}{\delta^D(X_i)^2}\left\{\frac{1-Z_i}{f(0|X_i)}-1\right\}p_1^Y(X_i)\\ =& -\frac{2Z_i-1}{f(Z|X)}\frac{\mathbb{E}[D|Z,X]}{\delta^D(X_i)^2}p_1^Y(X)\ .\end{aligned}$$

Therefore, we have that

$$\begin{aligned}\text{(S5.31)} =&\text{(S5.55)} + \text{(S5.56)} + \text{(S5.57)}\\ =&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\delta^D(X_i)-D_i\left\{\frac{Z_i}{f(1|X_i)}-\frac{1-Z_i}{f(0|X_i)}\right\}\right]\frac{p_1^Y(X_i)}{\delta^D(X_i)^2}\\ &+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbb{E}[D|Z=1,X_i]}{\delta^D(X_i)^2}\left\{\frac{Z_i}{f(1|X_i)}-1\right\}p_1^Y(X_i)\\ &-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbb{E}[D|Z=0,X_i]}{\delta^D(X_i)^2}\left\{\frac{1-Z_i}{f(0|X_i)}-1\right\}p_1^Y(X_i)+o_p(1)\\ =&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}D_i\cdot\frac{2Z_i-1}{f(Z_i|X_i)}\cdot\frac{p_1^Y(X_i)}{\delta^D(X_i)^2}+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{2Z_i-1}{\delta^D(X_i)^2}\cdot\frac{\mathbb{E}[D_i|Z_i,X_i]}{f(Z_i|X_i)}p_1^Y(X_i)+o_p(1)\ ,\end{aligned}$$

which justifies our Lemma 4.

S6 Proof of Theorem 3

By definition, $\hat{\theta} := (\hat{\lambda}_K, \hat{\beta}_K, \hat{\gamma}_K, \hat{\tau})$ solves the following equation:

$$\frac{1}{N}\sum_{i=1}^{N}g(Z_i,D_i,X_i,Y_i;\hat{\theta})=0.$$

Let $\mathbf{e}_{3K+1}$ be a $(3K+1)$-dimensional column vector whose last element is 1 and other components are all of 0's, and $\theta^* := (\lambda_K^*, \beta_K^*, \gamma_K^*, \tau)^\top$. By Taylor's Theorem we obtain

$$\begin{aligned}
&\sqrt{N}(\widehat{\tau} - \tau) = \sqrt{N} \cdot \mathbf{e}_{3K+1}^\top(\widehat{\theta} - \theta^*) \\
=&\mathbf{e}_{3K+1}^\top \left(\frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(Z_i, D_i, X_i, Y_i; \tilde{\theta})}{\partial \theta} \right)^{-1} \left(-\frac{1}{\sqrt{N}} \sum_{i=1}^{N} g(Z_i, D_i, X_i, Y_i; \theta^*) \right) \\
=& - \mathbf{e}_{3K+1}^\top \left(\mathbb{E}\left[\frac{\partial g(Z_i, D_i, X_i, Y_i; \theta^*)}{\partial \theta} \right] \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} g(Z_i, D_i, X_i, Y_i; \theta^*) + o_P(1),
\end{aligned} \tag{S6.75}$$

where and $\tilde{\theta}$ lies on the line joining θ^* and $\hat{\theta}$, and the last equality follows from the following claim:

$$\left\| \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(Z_i, D_i, X_i, Y_i; \tilde{\theta})}{\partial \theta} - \mathbb{E}\left[\frac{\partial g(Z, D, X, Y; \theta^*)}{\partial \theta} \right] \right\| \xrightarrow{P} o_p(1). \tag{S6.76}$$

Proof of Claim (S6.76). Since

$$\frac{\partial g(Z_i, D_i, X_i, Y_i; \tilde{\theta})}{\partial \theta} = \begin{pmatrix} G_{11} & G_{12} & G_{13} & G_{14} \\ G_{21} & G_{22} & G_{23} & G_{24} \\ G_{31} & G_{32} & G_{33} & G_{34} \\ G_{41} & G_{42} & G_{43} & G_{44} \end{pmatrix}$$

where

$$G_{11} = \frac{1}{N}\sum_{i=1}^{N} Z_i\rho''(\tilde{\lambda}_K^\top u_K(X_i))u_K(X_i)u_K^\top(X_i), G_{12} = 0_{K\times K}, G_{13} = 0_{K\times K}, G_{14} = 0_{K\times 1},$$

$$G_{21} = 0_{K\times K}, G_{22} = \frac{1}{N}\sum_{i=1}^{N}(1-Z_i)\rho''(\tilde{\beta}_K^\top u_K(X_i))u_K(X_i)u_K^\top(X_i), G_{23} = 0_{K\times K}, G_{24} = 0_{K\times 1},$$

$$G_{31} = \frac{1}{N}\sum_{i=1}^{N} D_iZ_i\rho''(\tilde{\lambda}_K^\top u_K(X_i))u_K(X_i)u_K^\top(X_i),$$

$$G_{32} = -\frac{1}{N}\sum_{i=1}^{N} D_i(1-Z_i)\rho''(\tilde{\beta}_K^\top u_K(X_i))u_K(X_i)u_K^\top(X_i),$$

$$G_{33} = -\frac{1}{N}\sum_{i=1}^{N} \rho_1''(\tilde{\gamma}_K^\top u_K(X_i))u_K(X_i)u_K^\top(X_i),$$

$$G_{34} = 0_{K\times 1},$$

$$G_{41} = \frac{1}{N}\sum_{i=1}^{N} Z_i\rho''(\tilde{\lambda}_K^\top u_K(X_i))Y_iu_K^\top(X_i)/\rho_1'(\tilde{\gamma}_K^\top u_K(X_i)),$$

$$G_{42} = -\frac{1}{N}\sum_{i=1}^{N}(1-Z_i)\rho''(\tilde{\beta}_K^\top u_K(X_i))Y_iu_K^\top(X_i)/\rho_1'(\tilde{\gamma}_K^\top u_K(X_i)),$$

$$G_{43} = -\frac{1}{N}\sum_{i=1}^{N}\frac{\{Z_i\rho'(\tilde{\lambda}_K^\top u_K(X_i)) - (1-Z_i)\rho'(\tilde{\beta}_K^\top u_K(X_i))\}\rho_1''(\tilde{\gamma}_K^\top u_K(X_i))Y_iu_K^\top(X_i)}{\rho_1'(\tilde{\gamma}_K^\top u_K(X_i))^2},$$

$$G_{44} = -1.$$

To establish (S6.76), it suffices to prove the upper left block is of convergence, i.e.,

$$\|G_{11} - \mathbb{E}\left[Z\rho''((\lambda_K^*)^\top u_K(X))u_K(X)u_K^\top(X)\right]\| \xrightarrow{p} 0 \ ,$$

and similar argument can be applied to show that other blocks are also of convergence. By Taylor's Theorem,

$$\begin{aligned}
&G_{11} - \mathbb{E}\left[Z\rho''((\lambda_K^*)^\top u_K(X))u_K(X)u_K^\top(X)\right] \\
=&\frac{1}{N}\sum_{i=1}^N Z_i\rho''(\lambda_K^*)^\top u_K(X_i))u_K(X_i)u_K^\top(X_i) - \mathbb{E}\left[Z\rho''((\lambda_K^*)^\top u_K(X))u_K(X)u_K^\top(X)\right] \qquad \text{(S6.77)} \\
&+ \frac{1}{N}\sum_{i=1}^N Z_i\rho''(\xi_3(X_i))u_K(X_i)u_K^\top(X_i)\cdot(\tilde{\lambda}_K - \lambda_K^*)^\top u_K(X_i)\ , \qquad \text{(S6.78)}
\end{aligned}$$

where $\xi_3(X_i)$ lies between $\tilde{\lambda}_K^\top u_K(X_i)$ and $(\lambda_K^*)^\top u_K(X_i)$. For the first term (S6.77), by computing its second moment, then using Chebyshev's inequality and Assumption 6, we can easily show (S6.77) is of $o_p(1)$. For the second term (S6.78), in light of the fact $\sup_{x\in\mathcal{X}}|\rho'''(\xi_3(x))| = O_p(1)$ and Lemma 2, we can deduce that

$$\begin{aligned}
|(\text{S6.78})| \leq&\frac{1}{N}\sum_{i=1}^N |\rho''(\xi_3(X_i))|\cdot\|u_K(X_i)\|^3\cdot\|\tilde{\lambda}_K - \lambda_K^*\| \\
\leq&\sup_{x\in\mathcal{X}}|\rho'''(\xi_3(x))|\cdot\|\hat{\lambda}_K - \lambda_K^*\|^2\cdot\frac{1}{N}\sum_{i=1}^N\|u_K(X_i)\|^3 \\
=&\sup_{x\in\mathcal{X}}|\rho'''(\xi_3(x))|\cdot\|\hat{\lambda}_K - \lambda_K^*\|^2\left\{\mathbb{E}[\|u_K(X)\|^3] + O_p\left(\zeta(K)^2\sqrt{\frac{K}{N}}\right)\right\} \\
\leq&\sup_{x\in\mathcal{X}}|\rho'''(\xi_3(x))|\cdot\|\hat{\lambda}_K - \lambda_K^*\|^2\left\{\zeta(K)\cdot\mathbb{E}[\|u_K(X)\|^2] + O_p\left(\zeta(K)^2\sqrt{\frac{K}{N}}\right)\right\}
\end{aligned}$$

$$\leq O_p(1) \cdot \frac{K}{N} \cdot \zeta(K) \cdot K = O_p\left(\zeta(K)\frac{K^2}{N}\right) = o_p(1) \ ,$$

Therefore, we obtain the claim (S6.76).

With (S6.75), the result $\sqrt{N}\{\widehat{\tau} - \tau\} \xrightarrow{d} N(0, V_{eff})$ of Theorem 2, and Slutsky's theorem, we obtain

$$\mathbf{e}_{3K+1}^{\top}\left(\mathbb{E}\left[\frac{\partial g(Z_i, D_i, X_i, Y_i; \theta^*)}{\partial \theta}\right]\right)^{-1} \frac{1}{\sqrt{N}}\sum_{i=1}^{N} g(Z_i, D_i, X_i, Y_i; \theta^*) \tag{S6.79}$$

$$= -\sqrt{N}\{\widehat{\tau} - \tau\} + o_P(1) \xrightarrow{d} N(0, V_{eff}).$$

Hence, the second moment of (S6.79) converges to the efficient variance, i.e.,

$$\mathbf{e}_{3K+1}^{\top}\left(\mathbb{E}\left[\frac{\partial g(D,Z,X,Y;\theta^*)}{\partial \theta}\right]\right)^{-1}\mathbb{E}[g(D,Z,X,Y;\theta^*)g^{\top}(D,Z,X,Y;\theta^*)]$$
$$\times\left(\mathbb{E}\left[\frac{\partial g(D,Z,X,Y;\theta^*)}{\partial \theta}\right]^{\top}\right)^{-1}\mathbf{e}_{3K+1} \to V_{eff}. \tag{S6.80}$$

The estimator for V_{eff} can be constructed by replacing $\mathbb{E}\left[\frac{\partial g(D,Z,X,Y;\theta^*)}{\partial \theta}\right] =:$ L and $\mathbb{E}[g(D,Z,X,Y;\theta^*)\ g^{\top}(D,Z,X,Y;\theta^*)] =: \Omega$ with their sample analogues:

$$\widehat{V} = \mathbf{e}_{3K+1}^{\top} \left\{ \widehat{L}^{-1} \cdot \widehat{\Omega} \cdot (\widehat{L}^{-1})^{\top} \right\} \mathbf{e}_{3K+1},$$

where

$$\widehat{L} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \theta} g(Z_i, D_i, X_i, Y_i; \hat{\theta}) \text{ and } \hat{\Omega} = \frac{1}{N} \sum_{i=1}^{N} g(Z_i, D_i, X_i, Y_i; \hat{\theta}) \cdot g(Z_i, D_i, X_i, Y_i; \hat{\theta})^{\top}.$$

To prove the consistency of the proposed variance estimator, it is sufficient to establish $\|\widehat{\Omega} - \Omega\| \xrightarrow{p} 0$ and $\|\widehat{L} - L\| \xrightarrow{p} 0$.

We shall elaborate the derivation for $\|\widehat{\Omega} - \Omega\| \xrightarrow{p} 0$, while $\|\widehat{L} - L\| \xrightarrow{p} 0$ can be similarly established. By applying the mean value theorem, we have

$$\widehat{\Omega} = \frac{1}{N} \sum_{i=1}^{N} g(D_i, Z_i, X_i, Y_i; \theta^*) g^{\top}(D_i, Z_i, X_i, Y_i; \theta^*) \tag{S6.81}$$

$$+ \frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(D_i, Z_i, X_i, Y_i; \tilde{\theta})}{\partial \theta} (\hat{\theta} - \theta^*)(\hat{\theta} - \theta^*)^{\top} \frac{\partial g^{\top}(D_i, Z_i, X_i, Y_i; \tilde{\theta})}{\partial \theta} \tag{S6.82}$$

$$+ \frac{2}{N} \sum_{i=1}^{N} g(D_i, Z_i, X_i, Y_i; \theta^*)(\hat{\theta} - \theta^*)^{\top} \frac{\partial g^{\top}(D_i, Z_i, X_i, Y_i; \tilde{\theta})}{\partial \theta}. \tag{S6.83}$$

where $\tilde{\theta}$ lies on the line joining $\hat{\theta}$ and θ^*. For the term (S6.81), computing its second moment and using the inequality $\{a+b+c+d\}^2 \leq 4a^2+4b^2+4c^2+4d^2$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N} g(D_i,Z_i,X_i,Y_i;\theta^*)g^\top(D_i,Z_i,X_i,Y_i;\theta^*) - \mathbb{E}\left[g(D,Z,X,Y;\theta^*)g^\top(D,Z,X,Y;\theta^*)\right]\right\|^2\right] \\
&\le \frac{1}{N}\mathbb{E}\left[\left\|g(D,Z,X,Y;\theta^*)g^\top(D,Z,X,Y;\theta^*)\right\|^2\right] \\
&= \frac{1}{N}\mathbb{E}\Bigg[\Bigg(g_1^\top(Z,X;\theta^*)g_1(Z,X;\theta^*) + g_2^\top(Z,X;\theta^*)g_2(Z,X;\theta^*) \\
&\qquad + g_3^\top(D,Z,X;\theta^*)g_3(D,Z,X;\theta^*) + g_4(D,Z,X,Y;\theta^*)^2\Bigg)^2\Bigg] \\
&\le \mathbb{E}\left[\frac{4}{N}\times (g_1^\top(Z,X;\theta^*)g_1(Z,X;\theta^*))^2\right] + \mathbb{E}\left[\frac{4}{N}\times (g_2^\top(Z,X;\theta^*)g_2(Z,X;\theta^*))^2\right] \\
&\quad + \mathbb{E}\left[\frac{4}{N}\times (g_3^\top(D,Z,X;\theta^*)g_3(D,Z,X;\theta^*))^2\right] + \mathbb{E}\left[\frac{4}{N}\times g_4(D,Z,X,Y;\theta^*)^4\right].
\end{aligned}
$$

Note

$$
\begin{aligned}
&\mathbb{E}\left[\frac{4}{N}\times (g_1^\top(Z,X;\theta^*)g_1(Z,X;\theta^*))^2\right] \\
&= \frac{4}{N}\times \mathbb{E}\left[Z^2(\rho'((\lambda_K^*)^\top u_K(X)) - 1)^2 (u_K^\top(X)u_K(X))^2\right] \\
&\le \frac{4}{N}\times O(\zeta^2(K))\times \mathbb{E}[u_K^\top(X)u_K(X)] \le O\left(\frac{\zeta^2(K)K}{N}\right).
\end{aligned}
$$

Similarly,

$$
\mathbb{E}\left[\frac{4}{N}\times (g_2^\top(Z,X;\theta^*)g_2(Z,X;\theta^*))^2\right] = O\left(\zeta^2(K)\frac{K}{N}\right),
$$

and

$$\mathbb{E}\left[\frac{4}{N}\times(g_3^\top(D,Z,X;\theta^*)g_3(D,Z,X;\theta^*))^2\right]=O\left(\zeta^2(K)\frac{K}{N}\right).$$

For the term $4\times N^{-1}\times\mathbb{E}\left[\{g_4(D,Z,X,Y;\theta^*)\}^4\right]$, we have

$$\begin{aligned}
&\mathbb{E}\left[\frac{4}{N}\times g_4(D,Z,X,Y;\theta^*)^4\right]\\
&=\frac{4}{N}\times\mathbb{E}\left[\left\{\frac{\{Z\rho'((\lambda_K^*)^\top u_K(X))-(1-Z)\rho'((\beta_K^*)^\top u_K(X))\}}{\rho_1'((\gamma_K^*)^\top u_K(X))}\cdot Y-\tau\right\}^4\right]\\
&\le\frac{32}{N}\times\mathbb{E}\left[\left\{\frac{\{Z\rho'((\lambda_K^*)^\top u_K(X))-(1-Z)\rho'((\beta_K^*)^\top u_K(X))\}}{\rho_1'((\gamma_K^*)^\top u_K(X))}\right\}^4\cdot Y^4+\tau^4\right]\\
&\le O\left(\frac{1}{N}\right)\times\mathbb{E}\left[\frac{Y^4}{\delta^D(X)^4}\right]+O\left(\frac{1}{N}\right)=O\left(\frac{1}{N}\right),
\end{aligned}$$

where the first inequality holds by using $(a+b)^p\le 2^{p-1}\cdot a^p+2^{p-1}\cdot b^p$ for $p\ge 1$; the second inequality follows from Lemma 1 that $\sup_{x\in\mathcal{X}}|\rho'\left((\lambda_K^*)^\top u_K(x)\right)|<\infty$, $\sup_{x\in\mathcal{X}}|\rho'\left((\beta_K^*)^\top u_K(x)\right)|<\infty$, and $\sup_{x\in\mathcal{X}}|\rho_1'\left((\gamma_K^*)^\top u_K(x)\right)-\delta^D(x)|=o(1)$; the last equality holds by using the assumption imposed in Theorem 3, i.e, $\mathbb{E}[Y^4/\delta_D^4(X)]<\infty$. Therefore, by Assumption 6, for the term (S6.81) we have (S6.81) $=\Omega+o_P(1)$. Similarly, it is easy to show that both (S6.82) and (S6.83) are of $o_P(1)$. With these results, we can conclude that $\widehat{\Omega}=\Omega+o_p(1)$.

S7 Proof of Theorem 4

Frölich (2007) derived the efficient influence function as:

$$\varphi_{LATE}(D,Z,X,Y)=\frac{1}{\mathbb{E}[\delta^D(X)]}\Bigg\{\frac{Z}{f(Z|X)}\Bigg(Y-\mathbb{E}[Y|Z=1,X]$$
$$-\tau_{LATE}\cdot\{D-\mathbb{E}[D|Z=1,X]\}\Bigg)-\frac{1-Z}{f(Z|X)}\Bigg(Y-\mathbb{E}[Y|Z=0,X]$$
$$-\tau_{LATE}\cdot\{D-\mathbb{E}[D|Z=0,X]\}\Bigg)+\delta^Y(X)-\tau_{LATE}\cdot\delta^D(X)\Bigg\}. \quad \text{(S7.84)}$$

It suffices to show that the influence function of $\sqrt{N}(\hat{\tau}_{LATE}-\tau_{LATE})$ equals to the efficient influence function above. Note that

$$\begin{aligned}
&\sqrt{N}(\hat{\tau}_{LATE}-\tau_{LATE})\\
=&\left[\frac{1}{N}\sum_{i=1}^N\widehat{d}_K(X_i)\right]^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N\{Z_i\hat{w}_K(1|X_i)Y_i-(1-Z_i)\hat{w}_K(0|X_i)Y_i\}-\sqrt{N}\cdot\tau_{LATE}\\
=&\left[\frac{1}{N}\sum_{i=1}^N\{Z_i\hat{w}_K(1|X_i)D_i-(1-Z_i)\hat{w}_K(0|X_i)D_i\}\right]^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N\{Z_i\hat{w}_K(1|X_i)Y_i-(1-Z_i)\hat{w}_K(0|X_i)Y_i\}-\sqrt{N}\cdot\tau_{LATE}\\
=&\left[\frac{1}{N}\sum_{i=1}^N\{Z_i\hat{w}_K(1|X_i)D_i-(1-Z_i)\hat{w}_K(0|X_i)D_i\}\right]^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N Z_i\hat{w}_K(1|X_i)Y_i-\sqrt{N}\cdot\mathbb{E}[\delta^D(X)]^{-1}\mathbb{E}[p_1^Y(X)]\\
&-\left[\frac{1}{N}\sum_{i=1}^N\{Z_i\hat{w}_K(1|X_i)D_i-(1-Z_i)\hat{w}_K(0|X_i)D_i\}\right]^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N(1-Z_i)\hat{w}_K(0|X_i)Y_i-\sqrt{N}\cdot\mathbb{E}[\delta^D(X)]^{-1}\mathbb{E}[p_0^Y(X)],
\end{aligned}$$

where the first equality holds in light of the balancing equation

$$\frac{1}{N}\sum_{i=1}^N\{Z_i\hat{w}_K(1|X_i)D_i-(1-Z_i)\hat{w}_K(0|X_i)D_i\}u_K(X_i)=\frac{1}{N}\sum_{i=1}^N\widehat{d}_K(X_i)u_K(X_i)$$

and the fact that $u_K(X)$ contains the constant 1.

We first derive the influence function of

$$\left[\frac{1}{N}\sum_{i=1}^{N}\{Z_i\hat{w}_K(1|X_i)D_i-(1-Z_i)\hat{w}_K(0|X_i)D_i\}\right]^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i\hat{w}_K(1|X_i)Y_i-\sqrt{N}\cdot\mathbb{E}[\delta^D(X)]^{-1}\mathbb{E}[p_1^Y(X)]\ ,$$

and similarly obtain that of

$$\left[\frac{1}{N}\sum_{i=1}^{N}\{Z_i\hat{w}_K(1|X_i)D_i-(1-Z_i)\hat{w}_K(0|X_i)D_i\}\right]^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-Z_i)\hat{w}_K(0|X_i)Y_i-\sqrt{N}\cdot\mathbb{E}[\delta^D(X)]^{-1}\mathbb{E}[p_0^Y(X)]\ .$$

Define $\delta_1(X) := \mathbb{E}[D|Z=1,X]$ and $\delta_0(X) = \mathbb{E}[D|Z=0,X]$, then $\delta^D(X) = \delta_1(X) - \delta_0(X)$. We also introduce the following notations:

$$\tilde{H}_K(X) = \tilde{\Phi}_K^\top\tilde{\Omega}_K^{-1}u_K(X)\ ,$$
$$\tilde{\Phi}_K = -\int_{\mathcal{X}} p_1^Y(x)f(1|x)\rho''(\tilde{\lambda}_K^\top u_K(x))u_K(x)dF_X(x)\ ,$$
$$\tilde{\Omega}_K = \frac{1}{N}\sum_{i=1}^{N} Z_i\rho''(\tilde{\lambda}_K^\top u_K(X_i))u_K(X_i)u_K^\top(X_i)\ ,$$

and

$$H_K(X) = \Phi_K^\top\Omega_K^{-1}u_K(X)\ ,$$
$$\Phi_K = -\int_{\mathcal{X}} p_1^Y(x)f(1|x)\rho''((\lambda_K^*)^\top u_K(x))u_K(x)dF_X(x)\ ,$$
$$\Omega_K = \mathbb{E}\left[f(1|X)\rho''((\lambda_K^*)^\top u_K(X))u_K(X)u_K^\top(X)\right]\ .$$

Note that $H_K(X)$ is the weighted L^2 projection of $-p_1^Y(X)$ on the space spanned by $u_K(X)$. Now, we can decompose

$$\left[\frac{1}{N}\sum_{i=1}^{N}\{Z_i\hat{w}_K(1|X_i)D_i-(1-Z_i)\hat{w}_K(0|X_i)D_i\}\right]^{-1}\sqrt{N}\sum_{i=1}^{N}Z_i\hat{w}_K(1|X_i)Y_i$$

as follows

$$\left[\frac{1}{N}\sum_{i=1}^{N}\{Z_i\hat{w}_K(1|X_i)D_i-(1-Z_i)\hat{w}_K(0|X_i)D_i\}\right]^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i\hat{w}_K(1|X_i)Y_i-\sqrt{N}\cdot\mathbb{E}[\delta^D(X)]^{-1}\mathbb{E}[p_1^Y(X)]$$

$$=\mathbb{E}\left[\delta^D(X)\right]^{-1}\cdot\sqrt{N}\left\{\frac{1}{N}\sum_{i=1}^{N}Z_i\hat{w}_K(1|X_i)Y_i-\mathbb{E}[p_1^Y(X)]\right\} \tag{S7.85}$$

$$-\mathbb{E}\left[\delta^D(X)\right]^{-1}\cdot\sqrt{N}\left\{\frac{1}{N}\sum_{i=1}^{N}\{Z_i\hat{w}_K(1|X_i)D_i-(1-Z_i)\hat{w}_K(0|X_i)D_i\}-\mathbb{E}\left[\delta^D(X)\right]\right\} \tag{S7.86}$$

$$\cdot\left[\frac{1}{N}\sum_{i=1}^{N}\{Z_i\hat{w}_K(1|X_i)D_i-(1-Z_i)\hat{w}_K(0|X_i)D_i\}\right]^{-1}\cdot\frac{1}{N}\sum_{i=1}^{N}Z_i\hat{w}_K(1|X_i)Y_i\ .$$

The following Lemma is key for establishing the asymptotic normality of $\hat{\tau}_{LATE}$. We omit the proof because it is similar to that of Theorem 2 of Chan et al. (2016).

Lemma 5.

$$\sqrt{N}\left\{\frac{1}{N}\sum_{i=1}^{N}Z_i\hat{w}_K(1|X_i)Y_i-\mathbb{E}[p_1^Y(X)]\right\}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{Z_iY_i}{f(1|X_i)}-p_1^Y(X_i)\left(\frac{Z_i}{f(1|X_i)}-1\right)-\mathbb{E}[p_1^Y(X)]\right\}+o_p(1)\ ,$$

$$\sqrt{N}\left\{\frac{1}{N}\sum_{i=1}^{N}(1-Z_i)\hat{w}_K(0|X_i)Y_i-\mathbb{E}[p_0^Y(X)]\right\}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{1-Z_i}{1-f(1|X_i)}Y_i-p_0^Y(X_i)\left(\frac{1-Z_i}{1-f(1|X_i)}-1\right)-\mathbb{E}[p_0^Y(X)]\right\}+o_p(1)\ ,$$

$$\sqrt{N}\left\{\frac{1}{N}\sum_{i=1}^{N}Z_i\hat{w}_K(1|X_i)D_i-\mathbb{E}[\delta_1(X)]\right\}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{Z_iD_i}{f(1|X_i)}-\delta_1(X_i)\left(\frac{Z_i}{f(1|X_i)}-1\right)-\mathbb{E}[\delta_1(X)]\right\}+o_p(1)\ ,$$

$$\sqrt{N}\left\{\frac{1}{N}\sum_{i=1}^{N}(1-Z_i)\hat{w}_K(0|X_i)D_i-\mathbb{E}[\delta_0(X)]\right\}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{1-Z_i}{1-f(1|X_i)}D_i-\delta_0(X_i)\left(\frac{1-Z_i}{1-f(1|X_i)}-1\right)-\mathbb{E}[\delta_0(X)]\right\}+o_p(1)\ .$$

Using Lemma 5, we can have that

$$\left[\frac{1}{N}\sum_{i=1}^{N}\{Z_i\hat{w}_K(1|X_i)D_i-(1-Z_i)\hat{w}_K(0|X_i)D_i\}\right]^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i\hat{w}_K(1|X_i)Y_i-\sqrt{N}\cdot\mathbb{E}[\delta^D(X)]^{-1}\mathbb{E}[p_1^Y(X)]$$

$$
\begin{aligned}
=&\mathbb{E}[\delta^D(X)]^{-1} \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{Z_i Y_i}{f(1|X_i)} - p_1^Y(X_i) \left(\frac{Z_i}{f(1|X_i)} - 1 \right) - \mathbb{E}[p_1^Y(X)] \right\} \\
&- \mathbb{E}[\delta^D(X)]^{-1} \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{Z_i D_i}{f(1|X_i)} - \frac{1-Z_i}{1-f(1|X_i)} \cdot D_i - \delta_1(X_i) \left(\frac{Z_i}{f(1|X_i)} - 1 \right) + \delta_0(X_i) \left(\frac{1-Z_i}{1-f(1|X_i)} - 1 \right) \right. \\
&\qquad \left. - \mathbb{E}[\delta^D(X)] \right\} \cdot \mathbb{E}[\delta^D(X)]^{-1} \cdot \mathbb{E}\left[\frac{Z}{f(1|X)} Y \right] + o_p(1) \\
=&\mathbb{E}[\delta^D(X)]^{-1} \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{Z_i}{f(1|X_i)} \left(Y_i - p_1^Y(X_i) \right) + p_1^Y(X_i) - \mathbb{E}[p_1^Y(X)] \right\} \\
&- \mathbb{E}[\delta^D(X)]^{-1} \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{Z_i}{f(1|X_i)} (D_i - \delta_1(X_i)) - \frac{1-Z_i}{1-f(1|X_i)} (D_i - \delta_0(X_i)) + \delta^D(X_i) - \mathbb{E}[\delta^D(X)] \right\} \\
&\cdot \mathbb{E}[\delta^D(X)]^{-1} \mathbb{E}[p_1^Y(X)]
\end{aligned}
$$

and

$$
\begin{aligned}
&\left[\frac{1}{N} \sum_{i=1}^N \{ Z_i \hat{w}_K(1|X_i) D_i - (1-Z_i) \hat{w}_K(0|X_i) D_i \} \right]^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N (1-Z_i) \hat{w}_K(0|X_i) Y_i - \sqrt{N} \cdot \mathbb{E}[\delta^D(X)]^{-1} \mathbb{E}[p_0^Y(X)] \\
=&\mathbb{E}[\delta^D(X)]^{-1} \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{1-Z_i}{1-f(1|X_i)} \left(Y_i - p_0^Y(X_i) \right) + p_0^Y(X_i) - \mathbb{E}[p_0^Y(X)] \right\} \\
&- \mathbb{E}[\delta^D(X)]^{-1} \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{Z_i}{f(1|X_i)} (D_i - \delta_1(X_i)) - \frac{1-Z_i}{1-f(1|X_i)} (D_i - \delta_0(X_i)) + \delta^D(X_i) - \mathbb{E}[\delta^D(X)] \right\} \\
&\cdot \mathbb{E}[\delta^D(X)]^{-1} \mathbb{E}[p_0^Y(X)] \ .
\end{aligned}
$$

Therefore, we have that

$$
\begin{aligned}
&\sqrt{N}(\hat{\tau}_{LATE} - \tau_{LATE}) \\
=&\mathbb{E}[\delta^D(X)]^{-1} \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{Z_i}{f(1|X_i)} \left(Y_i - p_1^Y(X_i) \right) - \frac{1-Z_i}{1-f(1|X_i)} \left(Y_i - p_0^Y(X_i) \right) + \delta^Y(X_i) - \mathbb{E}[\delta^Y(X)] \right\} \\
&- \mathbb{E}[\delta^D(X)]^{-1} \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{Z_i}{f(1|X_i)} (D_i - \delta_1(X_i)) - \frac{1-Z_i}{1-f(1|X_i)} (D_i - \delta_0(X_i)) + \delta^D(X_i) - \mathbb{E}[\delta^D(X)] \right\} \cdot \tau_{LATE} \\
=&\frac{1}{\sqrt{N}} \sum_{i=1}^N \varphi_{LATE}(D_i, Z_i, X_i, Y_i) \ .
\end{aligned}
$$

S8 Calibration Estimation with Estimated Design Weights

At the suggestion of the reviewer, in this section we show that the calibration estimator constructed from estimated initial design weights also produce efficient estimation of average treatment effects. For simplicity of presentation, we suppose there is no unmeasured confounder. Then the unconfoundedness condition holds, i.e. $D \perp \{Y(1), Y(0)\} | X$. The average treatment effect is identified by

$$\tau = \mathbb{E}\left[\frac{2D-1}{f(D|X)} Y\right], \tag{S8.87}$$

where $f(D|X)$ is the conditional probability mass function of D given X (the propensity score function). Suppose the propensity score function $f(D|X)$ is initially estimated via a parametric maximum likelihood, denoted by $f(D|X;\widehat{\iota})$, where $\widehat{\iota}$ is an estimated finite dimensional parameter. Let ι^* be the limit of $\widehat{\iota}$. Under regular conditions, the parametric MLE delivers $\sqrt{N}$-convergence rate, i.e. $\|\widehat{\iota} - \iota^*\| = O_P(N^{-1/2})$.

We consider minimizing the distance from the estimated sampling weights $f^{-1}(D|X;\widehat{\iota})$ under the distance measure defined by $D(a,b) := a\log(a/b) - a + b$:

$$\begin{cases} \min \sum_{i=1}^N \left\{ w_i \log\left(w_i \cdot f(D_i|X_i;\widehat{\iota})\right) - w_i + f^{-1}_{D|X}(1|X_i;\widehat{\iota}) \right\} \\ \text{subject to} \quad \frac{1}{N}\sum_{i=1}^N D_i w_i u_K(X_i) = \frac{1}{N}\sum_{i=1}^N u_K(X_i) = \frac{1}{N}\sum_{i=1}^N (1-D_i) w_i u_K(X_i). \end{cases} \tag{S8.88}$$

Applying the similar argument of Chan et al. (2016, Appendix B), the dual solution of (S8.88) is given by

$$\widehat{w}_K(D_i, X_i) := \frac{D_i}{f(D_i|X_i;\widehat{\iota})} \times \exp\left(-\widehat{\lambda}_K^\top u_K(X_i)\right) + \frac{1-D_i}{f(D_i|X_i;\widehat{\iota})} \times \exp\left(-\widehat{\beta}_K^\top u_K(X_i)\right),$$

where $\widehat{\lambda}_K, \widehat{\beta}_K \in \mathbb{R}^K$ maximize the following concave objective function:

$$\widehat{G}(\lambda,\beta) := \left\{-\frac{1}{N}\sum_{i=1}^N \frac{D_i}{f_{D|X}(D_i|X_i;\widehat{\theta})}\exp\left(-\lambda^\top u_K(X_i)\right) - \frac{1}{N}\sum_{i=1}^N \lambda^\top u_K(X_i)\right\} \tag{S8.89}$$
$$+\left\{-\frac{1}{N}\sum_{i=1}^N \frac{1-D_i}{f_{D|X}(D_i|X_i;\widehat{\theta})}\exp\left(-\beta^\top u_K(X_i)\right) - \frac{1}{N}\sum_{i=1}^N \beta^\top u_K(X_i)\right\}$$
$$=\widehat{G}_1(\lambda) + \widehat{G}_2(\beta), \tag{S8.90}$$

where the definitions of $\widehat{G}_1(\lambda)$ and $\widehat{G}_2(\beta)$ are obvious. The estimator of average treatment effects (S8.87) is given by

$$\widehat{\tau} = \frac{1}{N}\sum_{i=1}^N \left\{D_i \widehat{w}_K(1|X_i)Y_i - (1-D_i)\widehat{w}_K(0|X_i)Y_i\right\}.$$

Using a similar argument of proving Theorem 2, we can show that $\widehat{\tau}$ attains the efficiency bound of τ developed in Hahn (1998), i.e.

$$\sqrt{N}\{\widehat{\tau} - \tau\} = \frac{1}{\sqrt{N}}\sum_{i=1}^N \left\{\frac{D_i}{f(D_i|X_i)}Y_i - \left[\frac{D_i}{f(D_i|X_i)} - 1\right]\cdot p_1^Y(X_i) - \mathbb{E}[Y(1)]\right\} \tag{S8.91}$$
$$-\frac{1}{\sqrt{N}}\sum_{i=1}^N \left\{\frac{1-D_i}{f(D_i|X_i)}Y_i - \left[\frac{1-D_i}{f(D_i|X_i)} - 1\right]\cdot p_0^Y(X_i) - \mathbb{E}[Y(0)]\right\}$$

$$+ o_P(1),$$

where $p_d^Y(X) := \mathbb{E}\left[Y|X, D = d\right]$ for $d \in \{0, 1\}$.

Proof of (S8.91)**.** It is sufficient to prove

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{D_i\hat{w}_K(1|X_i)Y_i - \mathbb{E}[Y(1)]\} \\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{D_i}{f(D_i|X_i)}Y_i - \left[\frac{D_i}{f(D_i|X_i)} - 1\right]\cdot p_1^Y(X_i) - \mathbb{E}[Y(1)]\right\} + o_P(1).
\end{aligned} \tag{S8.92}$$

We introduce the following notation:

$$G_1^*(\lambda) = \mathbb{E}\left[-\frac{D}{f(D|X;\iota^*)}\exp\left(-\lambda^\top u_K(X)\right) - \lambda^\top u_K(X)\right],\ \lambda_K^* = \arg\max G_1^*(\lambda),$$

and

$$\begin{aligned}
\tilde{Q}_K(X) &= \tilde{\Psi}_K^\top \tilde{\Sigma}_K^{-1} u_K(X)\ , \\
\tilde{\Psi}_K &= \int_{\mathcal{X}} p_1^Y(x) f(D=1|x)\frac{\exp\left(-\tilde{\lambda}_K^\top u_K(x)\right)}{f(D=1|x;\tilde{\iota})} u_K(x) dF_X(x)\ , \\
\tilde{\Sigma}_K &= -\frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{f(D_i=1|X_i;\tilde{\iota})}\exp\left(-\widetilde{\lambda}_K^\top u_K(X_i)\right) u_K(X_i) u_K(X_i)^\top\ ,
\end{aligned}$$

and

$$Q_K(X) = \Psi_K^\top \Sigma_K^{-1} u_K(X) \ ,$$
$$\Psi_K = \int_{\mathcal{X}} p_1^Y(x) f(D=1|x) \frac{\exp\left(-(\lambda_K^*)^\top u_K(x)\right)}{f(D=1|x;\iota^*)} u_K(x) dF_X(x) \ ,$$
$$\Sigma_K = -\mathbb{E}\left[\frac{f(D=1|X)}{f(D=1|X;\iota^*)} \exp\left(-(\lambda_K^*)^\top u_K(X)\right) u_K(X) u_K(X)^\top\right] \ .$$

It should be noted that $Q_K(X)$ is the weighted L^2 projection of $-p_1^Y(X)$ on the space spanned by $u_K(X)$.

Using a similar decomposition of (S5.30)-(S5.38), we can decompose $N^{-1/2}\sum_{i=1}^N \{D_i \widehat{w}_K(1|X_i) Y_i - \mathbb{E}[Y(1)]\}$ as follows:

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \{D_i \hat{w}_K(1|X_i) Y_i - \mathbb{E}[Y(1)]\}$$
$$= \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ D_i \left(\hat{w}_K(1|X_i) - w_K^*(1|X_i)\right) Y_i - \int_{\mathcal{X}} p_1^Y(x) f(D=1|x) (\hat{w}_K(1|x) - w_K^*(1|x)) dF_X(x) \right\} \tag{S8.93}$$
$$+ \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \left(w_K^*(1|X_i) - \frac{1}{f(D=1|X_i)} \right) D_i Y_i - \mathbb{E}\left[p_1^Y(X) f(D=1|X) \left(w_K^*(1|X) - \frac{1}{f(D=1|X)} \right) \right] \right\} \tag{S8.94}$$
$$+ \sqrt{N} \cdot \mathbb{E}\left[p_1^Y(X) f(D=1|X) \left(w_K^*(1|X) - \frac{1}{f(D=1|X)} \right) \right] \tag{S8.95}$$
$$+ \Bigg[\sqrt{N} \int_{\mathcal{X}} p_1^Y(x) f(D=1|x) (\hat{w}_K(1|x) - w_K^*(1|x)) dF_X(x)$$
$$- \frac{1}{\sqrt{N}} \sum_{i=1}^N \left[\frac{D_i}{f_{D|X}(D_i|X_i;\iota^*)} \exp(-(\lambda_K^*)^\top u_K(X_i)) - 1 \right] \tilde{Q}_K(X_i)$$
$$+ \left\{ \frac{1}{N} \sum_{i=1}^N \widetilde{Q}_K(X_i) \cdot D_i \exp\left(-(\tilde{\lambda}_K)^\top u_K(X_i)\right) \frac{\partial_\iota f_{D|X}(D_i|X_i;\tilde{\iota})}{f_{D|X}^2(D_i|X_i;\tilde{\iota})} \right.$$
$$\left. + \int_{\mathcal{X}} p_1^Y(x) f(D=1|x) \exp\left(-\widetilde{\lambda}_K^\top u_K(x)\right) \frac{\partial_\iota f(D=1|x;\tilde{\iota})}{f^2(D=1|x;\tilde{\iota})} dF_X(x) \right\} \cdot \sqrt{N}\left\{\widehat{\iota} - \iota^*\right\} \Bigg] \tag{S8.96}$$

$$- \left\{ \frac{1}{N} \sum_{i=1}^{N} \widetilde{Q}_K(X_i) \cdot D_i \exp\left(-(\tilde{\lambda}_K)^\top u_K(X_i)\right) \frac{\partial_\iota f_{D|X}(D_i|X_i;\tilde{\iota})}{f^2_{D|X}(D_i|X_i;\tilde{\iota})} \right.$$
$$\left. + \int_{\mathcal{X}} p_1^Y(x) f(D=1|x) \exp\left(-\widetilde{\lambda}_K^\top u_K(x)\right) \frac{\partial_\iota f(D=1|x;\tilde{\iota})}{f^2(D=1|x;\tilde{\iota})} dF_X(x) \right\} \cdot \sqrt{N}\{\widehat{\iota} - \iota^*\} \tag{S8.97}$$
$$+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\frac{D_i}{f_{D|X}(D_i|X_i;\iota^*)} \exp(-(\lambda_K^*)^\top u_K(X_i)) - 1 \right] (\tilde{Q}_K(X_i) - Q_K(X_i)) \tag{S8.98}$$
$$+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left\{ \left[\frac{D_i}{f_{D|X}(D_i|X_i;\iota^*)} \exp(-(\lambda_K^*)^\top u_K(X_i)) - 1 \right] Q_K(X_i) + \left(\frac{D_i}{f(D_i=1|X_i)} - 1 \right) p_1^Y(X_i) \right\} \tag{S8.99}$$
$$- \left\{ \frac{1}{N} \sum_{i=1}^{N} \widetilde{Q}_K(X_i) \cdot D_i \exp\left(-(\tilde{\lambda}_K)^\top u_K(X_i)\right) \frac{\partial_\iota f_{D|X}(D_i|X_i;\tilde{\iota})}{f^2_{D|X}(D_i|X_i;\tilde{\iota})} \right.$$
$$\left. + \int_{\mathcal{X}} p_1^Y(x) f(D=1|x) \exp\left(-\widetilde{\lambda}_K^\top u_K(x)\right) \frac{\partial_\iota f(D=1|x;\tilde{\iota})}{f^2(D=1|x;\tilde{\iota})} dF_X(x) \right\} \cdot \sqrt{N}\{\widehat{\iota} - \iota^*\} \Bigg] \tag{S8.100}$$
$$+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left\{ \frac{D_i Y_i}{f(D_i=1|X_i)} - \left(\frac{D_i}{f(D_i=1|X_i)} - 1 \right) p_1^Y(X_i) - \mathbb{E}[Y(1)] \right\} . \tag{S8.101}$$

We claim that (S8.93)-(S8.100) are of $o_P(1)$, then (S8.92) holds true.

Using a similar argument of showing (S5.32)-(S5.37) are of $o_P(1)$, we can show that (S8.93)-(S8.99) are of $o_P(1)$. The term (S8.96) is exactly zero, which can be demonstrated as follows: using the first order condition, i.e. $\widehat{G}_1'(\widehat{\lambda}_K) = 0$, and mean value theorem, we have

$$\frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{D_i}{f_{D|X}(D_i|X_i;\iota^*)} \exp\left(-(\lambda_K^*)^\top u_K(X_i)\right) - 1 \right\} u_K(X_i)$$
$$- \frac{1}{N} \sum_{i=1}^{N} \frac{D_i}{f^2_{D|X}(D_i|X_i;\tilde{\iota})} \exp\left(-(\tilde{\lambda}_K)^\top u_K(X_i)\right) u_K(X_i) \cdot \partial_\iota f_{D|X}(D_i|X_i;\tilde{\iota}) \{\widehat{\iota} - \iota^*\}$$
$$= \left[\frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{D_i}{f_{D|X}(D_i|X_i;\tilde{\iota})} \exp\left(-(\tilde{\lambda}_K)^\top u_K(X_i)\right) \right\} u_K(X_i) u_K^\top(X_i) \right] \left\{ \widehat{\lambda}_K - \lambda_K^* \right\}.$$

Then

$$\widehat{\lambda}_K - \lambda_K^*$$

$$
\begin{aligned}
=&\left[\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{D_i}{f_{D|X}(D_i|X_i;\tilde{\iota})}\exp\left(-(\tilde{\lambda}_K)^{\top}u_K(X_i)\right)\right\}u_K(X_i)u_K^{\top}(X_i)\right]^{-1}\\
&\quad\times\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{D_i}{f_{D|X}(D_i|X_i;\iota^*)}\exp\left(-(\lambda_K^*)^{\top}u_K(X_i)\right)-1\right\}u_K(X_i)\\
&-\left[\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{D_i}{f_{D|X}(D_i|X_i;\tilde{\iota})}\exp\left(-(\tilde{\lambda}_K)^{\top}u_K(X_i)\right)\right\}u_K(X_i)u_K^{\top}(X_i)\right]^{-1}\\
&\quad\times\frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{f_{D|X}^2(D_i|X_i;\tilde{\iota})}\exp\left(-(\tilde{\lambda}_K)^{\top}u_K(X_i)\right)u_K(X_i)\cdot\partial_{\iota}f_{D|X}(D_i|X_i;\tilde{\iota})\left\{\widehat{\iota}-\iota^*\right\}\\
=&-\widetilde{\Sigma}_K\cdot\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)\left\{\frac{D_i}{f_{D|X}(D_i|X_i;\iota^*)}\exp\left(-(\lambda_K^*)^{\top}u_K(X_i)\right)-1\right\}\\
&+\widetilde{\Sigma}_K\cdot\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)\frac{D_i}{f_{D|X}^2(D_i|X_i;\tilde{\iota})}\exp\left(-(\tilde{\lambda}_K)^{\top}u_K(X_i)\right)\cdot\partial_{\iota}f_{D|X}(D_i|X_i;\tilde{\iota})\left\{\widehat{\iota}-\iota^*\right\}
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\sqrt{N}\int_{\mathcal{X}}p_1^Y(x)f(D=1|x)(\hat{w}_K(1|x)-w_K^*(1|x))dF_X(x)\\
=&\sqrt{N}\int_{\mathcal{X}}p_1^Y(x)f(D=1|x)\left(\frac{\exp\left(-\widehat{\lambda}_K^{\top}u_K(x)\right)}{f(D=1|x;\widehat{\iota})}-\frac{\exp\left(-(\lambda_K^*)^{\top}u_K(x)\right)}{f(D=1|x;\iota^*)}\right)dF_X(x)\\
=&-\sqrt{N}\int_{\mathcal{X}}p_1^Y(x)f(D=1|x)\exp\left(-\widetilde{\lambda}_K^{\top}u_K(x)\right)\frac{\partial_{\iota}f(D=1|x;\tilde{\iota})}{f^2(D=1|x;\widetilde{\iota})}dF_X(x)\cdot\left\{\widehat{\iota}-\iota^*\right\}\\
&-\sqrt{N}\int_{\mathcal{X}}p_1^Y(x)f(D=1|x)\frac{\exp\left(-\widetilde{\lambda}_K^{\top}u_K(x)\right)}{f(D=1|x;\tilde{\iota})}u_K^{\top}(x)dF_X(x)\cdot\left\{\widehat{\lambda}_K-\lambda_K^*\right\}\\
=&-\int_{\mathcal{X}}p_1^Y(x)f(D=1|x)\exp\left(-\widetilde{\lambda}_K^{\top}u_K(x)\right)\frac{\partial_{\iota}f(D=1|x;\tilde{\iota})}{f^2(D=1|x;\widetilde{\iota})}dF_X(x)\cdot\sqrt{N}\left\{\widehat{\iota}-\iota^*\right\}\\
&+\tilde{\Psi}_K^{\top}\cdot\widetilde{\Sigma}_K\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_K(X_i)\left\{\frac{D_i}{f_{D|X}(D_i|X_i;\iota^*)}\exp\left(-(\lambda_K^*)^{\top}u_K(X_i)\right)-1\right\}\\
&-\tilde{\Psi}_K^{\top}\cdot\widetilde{\Sigma}_K\cdot\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)D_i\frac{\partial_{\iota}f_{D|X}(D_i|X_i;\tilde{\iota})}{f_{D|X}^2(D_i|X_i;\tilde{\iota})}\exp\left(-(\tilde{\lambda}_K)^{\top}u_K(X_i)\right)\cdot\sqrt{N}\left\{\widehat{\iota}-\iota^*\right\}\\
=&-\left\{\frac{1}{N}\sum_{i=1}^{N}\widetilde{Q}_K(X_i)\cdot D_i\exp\left(-(\tilde{\lambda}_K)^{\top}u_K(X_i)\right)\frac{\partial_{\iota}f_{D|X}(D_i|X_i;\tilde{\iota})}{f_{D|X}^2(D_i|X_i;\tilde{\iota})}\right.\\
&\quad\left.+\int_{\mathcal{X}}p_1^Y(x)f(D=1|x)\exp\left(-\widetilde{\lambda}_K^{\top}u_K(x)\right)\frac{\partial_{\iota}f(D=1|x;\tilde{\iota})}{f^2(D=1|x;\widetilde{\iota})}dF_X(x)\right\}\cdot\sqrt{N}\left\{\widehat{\iota}-\iota^*\right\}\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{Q}_K(X_i)\left\{\frac{D_i}{f_{D|X}(D_i|X_i;\iota^*)}\exp\left(-(\lambda_K^*)^{\top}u_K(X_i)\right)-1\right\},
\end{aligned}
$$

which gives (S8.96) = 0.

The term (S8.100) is of $o_P(1)$ by noting $\sqrt{N}\{\widehat{\iota}-\iota^*\}=O_P(1)$ and

$$\left\{\frac{1}{N}\sum_{i=1}^{N}\widetilde{Q}_K(X_i)\cdot D_i\exp\left(-(\tilde{\lambda}_K)^\top u_K(X_i)\right)\frac{\partial_\iota f_{D|X}(D_i|X_i;\tilde{\iota})}{f^2_{D|X}(D_i|X_i;\tilde{\iota})}\right.$$
$$\left.+\int_{\mathcal{X}}p_1^Y(x)f(D=1|x)\exp\left(-\widetilde{\lambda}_K^\top u_K(x)\right)\frac{\partial_\iota f(D=1|x;\tilde{\iota})}{f^2(D=1|x;\widetilde{\iota})}dF_X(x)\right\}\xrightarrow{p}0,$$

because $\tilde{Q}_K(X)$ is a weighted least square projection of $p_1^Y(X)$ on $u_K(X)$.

Bibliography

Chan, K. C. G., S. C. P. Yam, and Z. Zhang (2016). Globally efficient nonparametric inference of average treatment effects by empirical balancing calibration weighting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(3), 673–700.

Donald, S. G., Y. Hsu, and R. P. Lieli (2014a). Testing the unconfoundedness assumption via inverse probability weighted estimators of (l)att. *Journal of Business & Economic Statistics 32*(3), 395–415.

Donald, S. G., Y. C. Hsu, and R. P. Lieli (2014b). Inverse probability weighted estimation of local average treatment effects: A higher order mse expansion. *Statistics & Probability Letters 95*(C), 132–138.

Frölich, M. (2007). Nonparametric iv estimation of local average treatment effects with covariates. *Journal of Econometrics 139*(1), 35–75.

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

Wang, L. and E. Tchetgen Tchetgen (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.