Supplementary Material to "Model Averaging for Estimating Treatment Effects"

Zhihao Zhao · Xinyu Zhang · Guohua Zou · Alan T.K. Wan · Geoffrey K.F. Tso

Appendix

A.1 Proof of Theorem 1

Let $\boldsymbol{b} = (b_1, \ldots, b_{n_2})'$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_{n_2})'$. Then

$$
\begin{aligned}
C_n(\boldsymbol{w}) &= \left\|\tilde{\boldsymbol{Y}} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2 + 2\sigma_t^2 tr\big(\boldsymbol{P}_{t0}(\boldsymbol{w})\big) + 2\sigma_c^2 tr\big(\boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\big) \\
&= \left\|\boldsymbol{\Delta} + \boldsymbol{b} + \boldsymbol{\eta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2 + 2\sigma_t^2 tr\big(\boldsymbol{P}_{t0}(\boldsymbol{w})\big) + 2\sigma_c^2 tr\big(\boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\big) \\
&= 2\left[\big\langle \boldsymbol{\Delta} + \boldsymbol{b} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w}), \boldsymbol{\eta}\big\rangle + \sigma_t^2 tr\big(\boldsymbol{P}_{t0}(\boldsymbol{w})\big) + \sigma_c^2 tr\big(\boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\big)\right] \\
&\quad + \left\|\boldsymbol{\Delta} + \boldsymbol{b} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2 + \|\boldsymbol{\eta}\|^2 \\
&:= \underline{A}(\boldsymbol{w}) + \underline{B}(\boldsymbol{w}) + \|\boldsymbol{\eta}\|^2. \qquad \text{(A.1)}
\end{aligned}
$$

Zhihao Zhao
School of Statistics, Capital University of Economics and Business, 121 Zhangjialukou, Huaxiang Fengtai District, Beijing 100070, China

✉ Xinyu Zhang
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 55 Zhongguancun East Road, Haidian District, Beijing 100190, China
E-mail: xinyu@amss.ac.cn

Guohua Zou
School of Mathematical Sciences, Capital Normal University, 105 West Third Ring Road North, Haidian District, Beijing 100048 China

Alan T.K. Wan
Department of Management Sciences, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Kowloon, Hong Kong, China

Geoffrey K.F. Tso
Department of Management Sciences, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Kowloon, Hong Kong, China

Let us first consider the second term on the right-hand-side of (A.1)

$$\underline{B}(\boldsymbol{w}) = \left\|\boldsymbol{\Delta} + \boldsymbol{b} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2 = \|\boldsymbol{b}\|^2 + 2\left\langle \boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w}), \boldsymbol{b}\right\rangle + \left\|\boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2.$$

From (11) and Condition (14), we have

$$\begin{aligned}
\frac{E\left(\|\boldsymbol{b}\|^2|\boldsymbol{U}\right)}{E\left(\|\boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\|^2|\boldsymbol{U}\right)} &\le E\left(\|\boldsymbol{b}\|^2\,|\boldsymbol{U}\right)/\xi_n \\
&= n_2\; O\left(\left(p\left(\frac{\log n}{n}\right)^{1/p}\right)^2\right)/\xi_n \\
&= O\left(\frac{n}{\log n}\right) O\left(p^2\left(\frac{\log n}{n}\right)^{2/p}\right)/\xi_n \\
&= O\left(p^2\left(\frac{n}{\log n}\right)^{1-(2/p)}\xi_n^{-1}\right) \\
&= o(1), \quad a.s. \qquad (A.2)
\end{aligned}$$

By (A.2) and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
& E\left(\left\langle \boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w}), \boldsymbol{b}\right\rangle|\boldsymbol{U}\right) \\
&\le E\left[\left(\left\|\boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2 \;\|\boldsymbol{b}\|^2\right)^{1/2}|\boldsymbol{U}\right] \\
&= E\left[\left(\left\|\boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\| \|\boldsymbol{b}\|\right)|\boldsymbol{U}\right] \\
&\le \left\{E\left(\left\|\boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2|\boldsymbol{U}\right) E\left(\|\boldsymbol{b}\|^2\,|\boldsymbol{U}\right)\right\}^{1/2} \\
&= E\left(\left\|\boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2|\boldsymbol{U}\right)\left\{\frac{E\left(\|\boldsymbol{b}\|^2\,|\boldsymbol{U}\right)}{E\left(\left\|\boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2|\boldsymbol{U}\right)}\right\}^{1/2} \\
&= R_n(\boldsymbol{w})\; o(1), \quad a.s. \qquad (A.3)
\end{aligned}$$

Then

$$\begin{aligned}
E(\underline{B}(\boldsymbol{w})|\boldsymbol{U}) &= E\left[\left\|\boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2 + 2\langle \boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w}), \boldsymbol{b}\rangle + \|\boldsymbol{b}\|^2|\boldsymbol{U}\right] \\
&\le R_n(\boldsymbol{w}) + E\left(\left\|\boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2|\boldsymbol{U}\right)\left\{\frac{E\left(\|\boldsymbol{b}\|^2\,|\boldsymbol{U}\right)}{E\left(\left\|\boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2|\boldsymbol{U}\right)}\right\}^{1/2}
\end{aligned}$$

$$
\begin{aligned}
&+E\left(\left\|\boldsymbol{\Delta}-\widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^{2}|\boldsymbol{U}\right)\frac{E\left(\|\boldsymbol{b}\|^{2}|\boldsymbol{U}\right)}{E\left(\left\|\boldsymbol{\Delta}-\widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^{2}|\boldsymbol{U}\right)}\\
&=R_n(\boldsymbol{w})+R_n(\boldsymbol{w})\ o(1)+R_n(\boldsymbol{w})\ o(1)\\
&=R_n(\boldsymbol{w})\left(1+o(1)\right),\quad a.s. \qquad (A.4)
\end{aligned}
$$

Let $\boldsymbol{\zeta}_{ta}=\boldsymbol{Y}_{ta}-\boldsymbol{f}_{ta}$, $\boldsymbol{v}_{ca}=\boldsymbol{Y}_{ca}-\boldsymbol{f}_{ca}$, $\widehat{\boldsymbol{f}}_t^{(k)}=\boldsymbol{P}_t^{(k)}\boldsymbol{Y}_{ta}$ and $\widehat{\boldsymbol{f}}_c^{(k)}=\boldsymbol{P}_{\tilde{c}}^{(k)}\boldsymbol{Y}_{ca}$. From (12), we obtain the model average estimator

$$
\begin{aligned}
\widehat{\boldsymbol{\Delta}}(\boldsymbol{w})&=\sum_{k=1}^{K}w_k\left(\widehat{\boldsymbol{f}}_t^{(k)}-\widehat{\boldsymbol{f}}_c^{(k)}\right)\\
&=\sum_{k=1}^{K}w_k\left[\boldsymbol{P}_t^{(k)}\left(\boldsymbol{f}_{ta}+\boldsymbol{\zeta}_{ta}\right)-\boldsymbol{P}_{\tilde{c}}^{(k)}\left(\boldsymbol{f}_{ca}+\boldsymbol{v}_{ca}\right)\right]\\
&=\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{f}_{ta}+\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}-\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{f}_{ca}-\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca},
\end{aligned}
$$

which yields

$$
\begin{aligned}
\boldsymbol{\Delta}+\boldsymbol{b}-\widehat{\boldsymbol{\Delta}}(\boldsymbol{w})&=\boldsymbol{f}_t-\boldsymbol{f}_c-\widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\\
&=\left(\boldsymbol{f}_t-\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{f}_{ta}\right)-\left(\boldsymbol{f}_c-\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{f}_{ca}\right)-\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}+\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}\\
&:=\boldsymbol{A}_t(\boldsymbol{w})-\boldsymbol{A}_{\tilde{c}}(\boldsymbol{w})-\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}+\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}. \qquad (A.5)
\end{aligned}
$$

Let $\boldsymbol{\zeta}_t=(\zeta_1^t,\ldots,\zeta_{n_2}^t)'$ and $\boldsymbol{v}_c=(v_1^t,\ldots,v_{n_2}^t)'$. From (3) and (4), we have

$$
E((\eta_m)^2|\boldsymbol{U})=\sigma_t^2+\sigma_c^2 \quad \text{and} \quad E(\boldsymbol{\zeta}_t'\boldsymbol{v}_c|\boldsymbol{U})=0. \qquad (A.6)
$$

Then from (A.5) and (A.6), we obtain

$$
\begin{aligned}
E\left(\underline{A}(\boldsymbol{w})|\boldsymbol{U}\right)&=2E\left[\left\{\left\langle\boldsymbol{A}_t(\boldsymbol{w})-\boldsymbol{A}_{\tilde{c}}(\boldsymbol{w})-\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}+\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca},\quad \boldsymbol{\zeta}_t-\boldsymbol{v}_c\right\rangle\right.\right.\\
&\quad\left.\left.+\ \sigma_t^2 tr\left(\boldsymbol{P}_{t0}(\boldsymbol{w})\right)+\sigma_c^2 tr\left(\boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\right)\right\}|\boldsymbol{U}\right]\\
&=2E\left\{\left\langle\boldsymbol{A}_t(\boldsymbol{w})-\boldsymbol{A}_{\tilde{c}}(\boldsymbol{w}),\boldsymbol{\eta}\right\rangle|\boldsymbol{U}\right\}+2E\left\{\left\langle-\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta},\boldsymbol{\zeta}_t\right\rangle|\boldsymbol{U}\right\}\\
&\quad+2\sigma_t^2 tr\left(\boldsymbol{P}_{t0}(\boldsymbol{w})\right)+2E\left(\left\langle\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca},-\boldsymbol{v}_c\right\rangle|\boldsymbol{U}\right)+2\sigma_t^2 tr\left(\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\right)\\
&\quad+2E\left(\left\langle-\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta},-\boldsymbol{v}_c\right\rangle|\boldsymbol{U}\right)+2E\left(\left\langle\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca},\boldsymbol{\zeta}_t\right\rangle|\boldsymbol{U}\right)\\
&=0 \qquad (A.7)
\end{aligned}
$$

and

$$
E(\|\boldsymbol{\eta}\|^2|\boldsymbol{U})=n_2(\sigma_t^2+\sigma_c^2). \qquad (A.8)
$$

Now, from (A.1), (A.4), (A.7) and (A.8), we have

$$
\begin{aligned}
E\left(C_n(\boldsymbol{w})|\boldsymbol{U}\right)&=E\left(\underline{A}(\boldsymbol{w})|\boldsymbol{U}\right)+E\left(\underline{B}(\boldsymbol{w})|\boldsymbol{U}\right)+E\left(\|\boldsymbol{\eta}\|^2|\boldsymbol{U}\right)\\
&=R_n(\boldsymbol{w})\left(1+o(1)\right)+n_2(\sigma_t^2+\sigma_c^2),\quad a.s. \qquad (A.9)
\end{aligned}
$$

This completes the proof of Theorem 1. □

A.2 Proof of Theorem 2

The proof of Theorem 2 is an application of Whittle (1960) and Chebyshev's inequality. Following Zhang et al. (2013) proof of their Theorem 2.1, we assume, for purposes of convenience, that $\boldsymbol{U}$ is non-stochastic. The proof that follows also applies to the case of stochastic $\boldsymbol{U}$ because all of the technical conditions imposed on $\boldsymbol{U}$ hold almost surely. In the following proof, we assume that $C_1, \ldots, C_{15}, C'_8, \ldots$, and C'_{13} are distinct constants.

Now, from (A.1), it follows that

$$\begin{aligned} C_n(\boldsymbol{w}) &= L_n(\boldsymbol{w}) + \underline{A}(\boldsymbol{w}) + (\underline{B}(\boldsymbol{w}) - L_n(\boldsymbol{w})) + \|\boldsymbol{\eta}\|^2 \\ &=: L_n(\boldsymbol{w}) + \underline{A}(\boldsymbol{w}) + \tilde{B}(\boldsymbol{w}) + \|\boldsymbol{\eta}\|^2. \end{aligned} \tag{A.10}$$

Theorem 2 is valid if the following hold

$$\sup_{\boldsymbol{w}\in H_n} \left|\underline{A}(\boldsymbol{w})\right|/R_n(\boldsymbol{w}) \to_p 0, \tag{A.11}$$

$$\sup_{\boldsymbol{w}\in H_n} \left|\tilde{B}(\boldsymbol{w})/R_n(\boldsymbol{w})\right| \to_p 0 \tag{A.12}$$

and

$$\sup_{\boldsymbol{w}\in H_n} \left|L_n(\boldsymbol{w})/R_n(\boldsymbol{w}) - 1\right| \to_p 0. \tag{A.13}$$

From $(A.5)$, we have

$$\begin{aligned} \underline{A}(\boldsymbol{w}) =& 2\left\langle \boldsymbol{\Delta} + \boldsymbol{b} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w}), \boldsymbol{\eta}\right\rangle + 2\sigma_t^2 tr\left(\boldsymbol{P}_t(\boldsymbol{w})\right) + 2\sigma_c^2 tr\left(\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\right) \\ =& 2\left\langle \boldsymbol{f}_t - \boldsymbol{f}_c - \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{f}_{ta} + \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{f}_{ca} - \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta} + \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}, \boldsymbol{\eta}\right\rangle \\ & + 2\sigma_t^2 tr\left(\boldsymbol{P}_{t0}(\boldsymbol{w})\right) + 2\sigma_c^2 tr\left(\boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\right) \\ =& 2\left\langle \boldsymbol{f}_t - \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{f}_{ta}, \boldsymbol{\eta}\right\rangle - 2\left\langle \boldsymbol{f}_c - \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{f}_{ca}, \boldsymbol{\eta}\right\rangle - 2\left\{\left\langle \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}, \boldsymbol{\eta}\right\rangle\right. \\ & \left. - \sigma_t^2 tr\left(\boldsymbol{P}_{t0}(\boldsymbol{w})\right)\right\} + 2\left\{\left\langle \boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\boldsymbol{v}_{ca}, \boldsymbol{\eta}\right\rangle + \sigma_c^2 tr\left(\boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\right)\right\} \\ =& 2\left\langle \boldsymbol{A}_t(\boldsymbol{w}), \boldsymbol{\eta}\right\rangle - 2\left\langle \boldsymbol{A}_{\tilde{c}}(\boldsymbol{w}), \boldsymbol{\eta}\right\rangle - 2\left\{\left\langle \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}, \boldsymbol{\eta}\right\rangle - \sigma_t^2 tr\left(\boldsymbol{P}_{t0}(\boldsymbol{w})\right)\right\} \\ & + 2\left\{\left\langle \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}, \boldsymbol{\eta}\right\rangle + \sigma_c^2 tr\left(\boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\right)\right\}. \end{aligned} \tag{A.14}$$

Hence (A.3) is valid if the following conditions hold

$$\sup_{\boldsymbol{w}\in H_n} \left|\left\langle \boldsymbol{A}_t(\boldsymbol{w}), \boldsymbol{\eta}\right\rangle\right| / R_n(\boldsymbol{w}) \to_p 0, \tag{A.15}$$

$$\sup_{\boldsymbol{w}\in H_n} \left|\left\langle \boldsymbol{A}_{\tilde{c}}(\boldsymbol{w}), \boldsymbol{\eta}\right\rangle\right| / R_n(\boldsymbol{w}) \to_p 0, \tag{A.16}$$

$$\sup_{\boldsymbol{w}\in H_n} \left|\left\langle \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}, \boldsymbol{\eta}\right\rangle - \sigma_t^2 tr\left(\boldsymbol{P}_{t0}(\boldsymbol{w})\right)\right| / R_n(\boldsymbol{w}) \to_p 0 \tag{A.17}$$

and

$$\sup_{\boldsymbol{w}\in H_n} \left|\left\langle \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}, \boldsymbol{\eta}\right\rangle + \sigma_c^2 tr\big(\boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\big)\right| / R_n(\boldsymbol{w}) \to_p 0. \qquad \text{(A.18)}$$

Note that Condition (17) implies that

$$E\left((\zeta_m^t)^{4G}|\boldsymbol{U}\right) \le \kappa < \infty \text{ and } E\left((v_m^c)^{4G}|\boldsymbol{U}\right) \le \kappa < \infty, m = 1, \ldots, n_2. \qquad \text{(A.19)}$$

To prove (A.15), using (18), (A.19), Chebyshev's inequality and Theorem 2 of Whittle (1960), and following steps similar to the proof of Equation (A.1) in Wan et al. (2010), we have, for any $\delta > 0$,

$$\begin{aligned}
& P\left\{\sup_{\boldsymbol{w}\in H_n} \left|\left\langle \boldsymbol{A}_t(\boldsymbol{w}), \boldsymbol{\eta}\right\rangle\right| / R_n(\boldsymbol{w}) > \delta\right\} \\
\le\ & P\left\{\sup_{\boldsymbol{w}\in H_n} \sum_{k=1}^{K} w_k \left|\boldsymbol{\eta}'\left(\boldsymbol{f}_t - \boldsymbol{P}_t^{(k)}\boldsymbol{f}_{ta}\right)\right| > \delta\xi_n\right\} \\
\le\ & P\left\{\max_{1\le k\le K} \left|\boldsymbol{\eta}'\left(\boldsymbol{f}_t - \boldsymbol{P}_t^{(k)}\boldsymbol{f}_{ta}\right)\right| > \delta\xi_n\right\} \\
=\ & P\left\{\left\{\left|\left\langle \boldsymbol{\eta}, \boldsymbol{A}_t(\boldsymbol{w}_1^0)\right\rangle\right| > \delta\xi_n\right\} \cup \left\{\left|\left\langle \boldsymbol{\eta}, \boldsymbol{A}_t(\boldsymbol{w}_2^0)\right\rangle\right| > \delta\xi_n\right\}\right. \\
& \quad \left. \cup \cdots \cup \left\{\left|\left\langle \boldsymbol{\eta}, \boldsymbol{A}_t(\boldsymbol{w}_k^0)\right\rangle\right| > \delta\xi_n\right\}\right\} \\
\le\ & \sum_{k=1}^{K} P\left\{\left|\left\langle \boldsymbol{\eta}, \boldsymbol{A}_t(\boldsymbol{w}_k^0)\right\rangle\right| > \delta\xi_n\right\} \\
\le\ & \sum_{k=1}^{K} E\left\{\frac{\left\langle \boldsymbol{\eta}, \boldsymbol{A}_t(\boldsymbol{w}_k^0)\right\rangle^{2G}}{\delta^{2G}\xi_n^{2G}}\right\} \\
\le\ & C_1\delta^{-2G}\xi_n^{-2G} \sum_{k=1}^{K} \left\|\boldsymbol{A}_t(\boldsymbol{w}_k^0)\right\|^{2G} \to 0.
\end{aligned}$$

Similarly, for (A.16), we obtain

$$P\left\{\sup_{\boldsymbol{w}\in H_n} \left|\left\langle \boldsymbol{A}_{\tilde{c}}(\boldsymbol{w}), \boldsymbol{\eta}\right\rangle\right| / R_n(\boldsymbol{w}) > \delta\right\} \le C_2\delta^{-2G}\xi_n^{-2G} \sum_{k=1}^{K} \left\|\boldsymbol{A}_{\tilde{c}}(\boldsymbol{w}_k^0)\right\|^{2G} \to 0.$$

Now, let us prove (A.17). As

$$\begin{aligned}
& P\left\{\sup_{\boldsymbol{w}\in H_n} \left|\left\langle \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}, \boldsymbol{\eta}\right\rangle - \sigma_t^2 tr\big(\boldsymbol{P}_{t0}(\boldsymbol{w})\big)\right| / R_n(\boldsymbol{w}) > \delta\right\} \\
=\ & P\left\{\sup_{\boldsymbol{w}\in H_n} \left|\left\langle \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}, \boldsymbol{\zeta}_t - \boldsymbol{v}_c\right\rangle - \sigma_t^2 tr\big(\boldsymbol{P}_{t0}(\boldsymbol{w})\big)\right| / R_n(\boldsymbol{w}) > \delta\right\} \\
\le\ & P\left\{\sup_{\boldsymbol{w}\in H_n} \left|\left\langle \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}, \boldsymbol{\zeta}_t\right\rangle - \sigma_t^2 tr\big(\boldsymbol{P}_{t0}(\boldsymbol{w})\big)\right| / R_n(\boldsymbol{w}) > \delta\right\} \\
& + P\left\{\sup_{\boldsymbol{w}\in H_n} \left|\left\langle \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}, \boldsymbol{v}_c\right\rangle\right| / R_n(\boldsymbol{w}) > \delta\right\}, \qquad \text{(A.20)}
\end{aligned}$$

it suffices to show that the two terms on the right hand side of (A.20) approach zero. Note that the first term is bounded by

$$
\begin{aligned}
& P\left\{\sup_{\boldsymbol{w}\in H_n}\left|\left\langle \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta},\boldsymbol{\zeta}_t\right\rangle-\sigma_t^2 tr\left(\boldsymbol{P}_{t0}(\boldsymbol{w})\right)\right|/R_n(\boldsymbol{w})>\delta\right\}\\
\leq & \sum_{k=1}^{K} P\left\{\left|\left\langle \boldsymbol{P}_t(\boldsymbol{w}_k^0)\boldsymbol{\zeta}_{ta},\boldsymbol{\zeta}_t\right\rangle-\sigma_t^2 tr\left(\boldsymbol{P}_{t0}(\boldsymbol{w}_k^0)\right)\right|>\delta\xi_n\right\}\\
\leq & \sum_{k=1}^{K} P\left\{\left|\left\langle \boldsymbol{P}_{t0}(\boldsymbol{w}_k^0)\boldsymbol{\zeta}_{ta},\boldsymbol{\zeta}_{ta}\right\rangle-\sigma_t^2 tr\left(\boldsymbol{P}_{t0}(\boldsymbol{w}_k^0)\right)\right|>\delta\xi_n\right\}\\
\leq & \sum_{k=1}^{K} E\left[\frac{\left\{\left\langle \boldsymbol{P}_{t0}(\boldsymbol{w}_k^0)\boldsymbol{\zeta}_{ta},\boldsymbol{\zeta}_{ta}\right\rangle-\sigma_t^2 tr\left(\boldsymbol{P}_{t0}(\boldsymbol{w}_k^0)\right)\right\}^{2G}}{\delta^{2G}\xi_n^{2G}}\right]\\
\leq & \; C_3\delta^{-2G}\xi_n^{-2G}\sum_{k=1}^{K}\left[tr\left(\boldsymbol{P}_{t0}(\boldsymbol{w}_k^0)\boldsymbol{P}_{t0}'(\boldsymbol{w}_k^0)\right)\right]^G\\
= & \; C_3\delta^{-2G}\xi_n^{-2G}\sum_{k=1}^{K}\left[tr\left(\boldsymbol{P}_{t}(\boldsymbol{w}_k^0)\boldsymbol{P}_{t}'(\boldsymbol{w}_k^0)\right)\right]^G\\
\to & \; 0,
\end{aligned}
$$

while the second term is given by

$$
\begin{aligned}
& P\left\{\sup_{\boldsymbol{w}\in H_n}\left|\left\langle \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta},\boldsymbol{v}_c\right\rangle\right|/R_n(\boldsymbol{w})>\delta\right\}\\
\leq & \sum_{k=1}^{K} P\left\{\left|\left\langle \boldsymbol{P}_t(\boldsymbol{w}_k^0)\boldsymbol{\zeta}_{ta},\boldsymbol{v}_c\right\rangle-E\left\langle \boldsymbol{P}_t(\boldsymbol{w}_k^0)\boldsymbol{\zeta}_{ta},\boldsymbol{v}_c\right\rangle\right|>\delta\xi_n\right\}\\
= & \sum_{k=1}^{K} P\left[\left|\left\langle \boldsymbol{P}_t(\boldsymbol{w}_k^0)\boldsymbol{\zeta}_{ta},\boldsymbol{v}_{ca}\right\rangle\right|>\delta\xi_n\right]\\
\leq & \sum_{k=1}^{K} E\left[\frac{\left\{\left\langle \boldsymbol{P}_t(\boldsymbol{w}_k^0)\boldsymbol{\zeta}_{ta},\boldsymbol{v}_c\right\rangle\right\}^2}{\delta^2\xi_n^2}\right]\\
\leq & \sum_{k=1}^{K} E\left[\frac{\left\{\left\langle \boldsymbol{\zeta}_{ta}'\boldsymbol{P}_t'(\boldsymbol{w}_k^0)\boldsymbol{v}_c\boldsymbol{v}_c'\boldsymbol{P}_t(\boldsymbol{w}_k^0)\boldsymbol{\zeta}_{ta}\right\rangle\right\}}{\delta^2\xi_n^2}\right]\\
\leq & \; C_4\delta^{-2}\xi_n^{-2}\sum_{k=1}^{K}\left[tr\left(\boldsymbol{P}_t(\boldsymbol{w}_k^0)\boldsymbol{P}_t'(\boldsymbol{w}_k^0)\right)\right]\\
\to & \; 0. \qquad\qquad (A.21)
\end{aligned}
$$

Hence (A.17) is valid.

Similarly, for (A.18), it is observed that

$$
P\left\{\sup_{\boldsymbol{w}\in H_n}\left|\left\langle \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca},\boldsymbol{\eta}\right\rangle+\sigma_c^2 tr\left(\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\right)\right|/R_n(\boldsymbol{w})>\delta\right\}
$$

$$
\begin{aligned}
&= P\left\{\sup_{\boldsymbol{w}\in H_n}\left|\left\langle \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}, \boldsymbol{\zeta}_t - \boldsymbol{v}_c\right\rangle + \sigma_c^2 tr\big(\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\big)\right|/R_n(\boldsymbol{w}) > \delta\right\} \\
&\le P\left\{\sup_{\boldsymbol{w}\in H_n}\left|\left\langle \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}, \boldsymbol{v}_c\right\rangle - \sigma_c^2 tr\big(\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\big)\right|/R_n(\boldsymbol{w}) > \delta\right\} \\
&\quad + P\left\{\sup_{\boldsymbol{w}\in H_n}\left|\left\langle \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}, \boldsymbol{\zeta}_t\right\rangle\right|/R_n(\boldsymbol{w}) > \delta\right\} \\
&\le C_5\delta^{-2G}\xi_n^{-2G}\sum_{k=1}^{K}\left[tr\big(\boldsymbol{P}_{\tilde{c}}'(\boldsymbol{w}_k^0)\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w}_k^0)\big)\right]^G + C_6\delta^{-2}\xi_n^{-2}\sum_{k=1}^{K}\left[tr\big(\boldsymbol{P}_{\tilde{c}}'(\boldsymbol{w}_k^0)\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w}_k^0)\big)\right] \\
&\to 0. \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (A.22)
\end{aligned}
$$

Given (A.15)-(A.18), it is straightforward to see that (A.3) is valid.

Now, to prove (A.12), let us first consider

$$
\begin{aligned}
\underline{B}(\boldsymbol{w}) &= \left\|\boldsymbol{\Delta} + \boldsymbol{b} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2 \\
&= \left\|\boldsymbol{A}_t(\boldsymbol{w}) - \boldsymbol{A}_{\tilde{c}}(\boldsymbol{w})\right\|^2 + \left\|\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta} - \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}\right\|^2 \\
&\quad -2\left\langle \boldsymbol{A}_t(\boldsymbol{w}) - \boldsymbol{A}_{\tilde{c}}(\boldsymbol{w}), \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta} - \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}\right\rangle. \qquad (A.23)
\end{aligned}
$$

Define

$$
\begin{aligned}
\tilde{R}_n(\boldsymbol{w}) := E\big(\underline{B}(\boldsymbol{w})\big|\boldsymbol{U}\big) &= E\Big[\left\|\boldsymbol{A}_t(\boldsymbol{w}) - \boldsymbol{A}_{\tilde{c}}(\boldsymbol{w})\right\|^2 + \left\|\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta} - \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}\right\|^2 \\
&\quad -2\left\langle \boldsymbol{A}_t(\boldsymbol{w}) - \boldsymbol{A}_{\tilde{c}}(\boldsymbol{w}), \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta} - \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}\right\rangle\Big|\boldsymbol{U}\Big] \\
&= \left\|\boldsymbol{A}_t(\boldsymbol{w}) - \boldsymbol{A}_{\tilde{c}}(\boldsymbol{w})\right\|^2 + \sigma_t^2 tr\big(\boldsymbol{P}_{\tilde{c}}'(\boldsymbol{w})\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\big) \\
&\quad + \sigma_t^2 tr\big(\boldsymbol{P}_t'(\boldsymbol{w})\boldsymbol{P}_t(\boldsymbol{w})\big). \qquad (A.24)
\end{aligned}
$$

Then from (A.2) and (A.3), we obtain

$$
\begin{aligned}
&\sup_{\boldsymbol{w}\in H_n}\left|\tilde{R}_n(\boldsymbol{w}) - R_n(\boldsymbol{w})\right|/R_n(\boldsymbol{w}) \\
&= \sup_{\boldsymbol{w}\in H_n} E\left[\left(\left\|\boldsymbol{\Delta} + \boldsymbol{b} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2 - \left\|\boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2\right)\Big|\boldsymbol{U}\right]/R_n(\boldsymbol{w}) \\
&= \sup_{\boldsymbol{w}\in H_n} E\left[2\left(\left\langle \boldsymbol{\Delta} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{w}), \boldsymbol{b}\right\rangle + \|\boldsymbol{b}\|^2\right)\Big|\boldsymbol{U}\right]/R_n(\boldsymbol{w}) \\
&= o(1). \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (A.25)
\end{aligned}
$$

By (A.23)-(A.25), we have

$$
\begin{aligned}
&\sup_{\boldsymbol{w}\in H_n}\left|\underline{B}(\boldsymbol{w})/R_n(\boldsymbol{w}) - 1\right| \\
&= \sup_{\boldsymbol{w}\in H_n}\left[\left\{\underline{B}(\boldsymbol{w}) - \tilde{R}_n(\boldsymbol{w})\right\} + \left\{\tilde{R}_n(\boldsymbol{w}) - R_n(\boldsymbol{w})\right\}\right]/R_n(\boldsymbol{w}) \\
&= \sup_{\boldsymbol{w}\in H_n}\Big\{\left\|\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta} - \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}\right\|^2 - \sigma_c^2 tr\big(\boldsymbol{P}_{\tilde{c}}'(\boldsymbol{w})\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\big) - \sigma_t^2 tr\big(\boldsymbol{P}_t'(\boldsymbol{w})\boldsymbol{P}_t(\boldsymbol{w})\big)
\end{aligned}
$$

$$
\begin{aligned}
&-2\Big\langle \boldsymbol{A}_t(\boldsymbol{w})-\boldsymbol{A}_{\tilde{c}}(\boldsymbol{w}),\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}-\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}\Big\rangle\Big\}/R_n(\boldsymbol{w})+o(1)\\
=&\sup_{\boldsymbol{w}\in H_n}\Big[\Big\{\|\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}\|^2-\sigma_t^2 tr\big(\boldsymbol{P}_t'(\boldsymbol{w})\boldsymbol{P}_t(\boldsymbol{w})\big)\Big\}-2\Big\langle \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta},\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}\Big\rangle\\
&+\Big\{\|\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}\|^2-\sigma_c^2 tr\big(\boldsymbol{P}_{\tilde{c}}'(\boldsymbol{w})\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\big)\Big\}-2\Big\langle \boldsymbol{A}_t(\boldsymbol{w}),\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}\Big\rangle\\
&+2\Big\langle \boldsymbol{A}_t(\boldsymbol{w}),\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca})\Big\rangle+2\Big\langle \boldsymbol{A}_{\tilde{c}}(\boldsymbol{w}),\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}\Big\rangle\\
&-2\Big\langle \boldsymbol{A}_{\tilde{c}}(\boldsymbol{w}),\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}\Big\rangle\Big]/R_n(\boldsymbol{w})+o(1). \qquad \text{(A.26)}
\end{aligned}
$$

Next, we will show that the terms on the right hand side of (A.26) converge to zero in probability. Denote the maximum singular value by $\bar{\lambda}$. Following Wan et al. (2010), we have

$$
\begin{aligned}
\bar{\lambda}\big(\boldsymbol{P}_t^{(k)}\big)&=\bar{\lambda}\big(\boldsymbol{U}_t^{(k)}(\boldsymbol{U}_{ta}^{(k)\prime}\boldsymbol{U}_{ta}^{(k)})^{-1}\boldsymbol{U}_{ta}^{(k)\prime}\big)\\
&=\bar{\lambda}^{1/2}\big(\boldsymbol{U}_t^{(k)}(\boldsymbol{U}_{ta}^{(k)\prime}\boldsymbol{U}_{ta}^{(k)})^{-1}\boldsymbol{U}_{ta}^{(k)\prime}\boldsymbol{U}_{ta}^{(k)}(\boldsymbol{U}_{ta}^{(k)\prime}\boldsymbol{U}_{ta}^{(k)})^{-1}\boldsymbol{U}_t^{(k)\prime}\big)\\
&=\bar{\lambda}\big(\boldsymbol{U}_t^{(k)}(\boldsymbol{U}_{ta}^{(k)\prime}\boldsymbol{U}_{ta}^{(k)})^{-1}\boldsymbol{U}_t^{(k)\prime}\big)\\
&=\bar{\lambda}\big(\boldsymbol{U}_t^{(k)}(\boldsymbol{U}_t^{(k)\prime}\boldsymbol{U}_t^{(k)}+\boldsymbol{F})^{-1}\boldsymbol{U}_t^{(k)\prime}\big)\quad(\exists\boldsymbol{F}\geq 0)\\
&\leq 1.
\end{aligned}
$$

Recognising that $\bar{\lambda}\left(\boldsymbol{P}_t(\boldsymbol{w}^*)\right)\leq\max_{1\leq k\leq K}\bar{\lambda}\left(\boldsymbol{P}_t^{(k)}\right)\leq 1$ for any $\boldsymbol{w},\boldsymbol{w}^*\in H_n$, we have

$$
\begin{aligned}
tr\{\boldsymbol{P}_t'(\boldsymbol{w})\boldsymbol{P}_t(\boldsymbol{w}^*)\boldsymbol{P}_t'(\boldsymbol{w}^*)\boldsymbol{P}_t(\boldsymbol{w})\}&\leq\bar{\lambda}^2\big(\boldsymbol{P}_t(\boldsymbol{w}^*)\big)tr\{\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{P}_t'(\boldsymbol{w})\}\\
&\leq tr\{\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{P}_t'(\boldsymbol{w})\}.
\end{aligned}
$$

Furthermore, note that

$$
\begin{aligned}
&\bar{\lambda}\big(\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w}^*)\big)\leq\bar{\lambda}\big(\boldsymbol{P}_c(\boldsymbol{w}^*)\big)+\bar{\lambda}\big(\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w}^*)-\boldsymbol{P}_c(\boldsymbol{w}^*)\big)\\
&\leq 1+\max_{1\leq k\leq K}\bar{\lambda}^{1/2}\Big[\Big\{(\boldsymbol{U}_t^{(k)}-\boldsymbol{U}_c^{(k)})(\boldsymbol{U}_{ca}^{(k)\prime}\boldsymbol{U}_{ca}^{(k)})^{-1}\boldsymbol{U}_{ca}^{(k)\prime}\Big\}\\
&\qquad\qquad\Big\{(\boldsymbol{U}_t^{(k)}-\boldsymbol{U}_c^{(k)})(\boldsymbol{U}_{ca}^{(k)\prime}\boldsymbol{U}_{ca}^{(k)})^{-1}\boldsymbol{U}_{ca}^{(k)\prime}\Big\}'\Big]\\
&=1+\max_{k=1,\ldots,K}\bar{\lambda}^{1/2}\Big\{(\boldsymbol{U}_t^{(k)}-\boldsymbol{U}_c^{(k)})(\boldsymbol{U}_{ca}^{(k)\prime}\boldsymbol{U}_{ca}^{(k)})^{-1}(\boldsymbol{U}_t^{(k)}-\boldsymbol{U}_c^{(k)})'\Big\}\\
&\leq 1+\max_{1\leq k\leq K}\bar{\lambda}^{1/2}\left\{\left(n_c^{-1}\boldsymbol{U}_{ca}^{(k)\prime}\boldsymbol{U}_{ca}^{(k)}\right)^{-1}\right\}\\
&\quad\times\max_{1\leq k\leq K}\bar{\lambda}^{1/2}\left\{n_c^{-1}\left(\boldsymbol{U}_t^{(k)}-\boldsymbol{U}_c^{(k)}\right)'\left(\boldsymbol{U}_t^{(k)}-\boldsymbol{U}_c^{(k)}\right)\right\}\\
&\leq 1+C_0\max_{1\leq k\leq K}\bar{\lambda}^{1/2}\left\{n_c^{-1}\left(\boldsymbol{U}_t^{(k)}-\boldsymbol{U}_c^{(k)}\right)'\left(\boldsymbol{U}_t^{(k)}-\boldsymbol{U}_c^{(k)}\right)\right\}\\
&\leq 1+C_0\max_{1\leq k\leq K}tr^{1/2}\left\{n_c^{-1}\left(\boldsymbol{U}_t^{(k)}-\boldsymbol{U}_c^{(k)}\right)'\left(\boldsymbol{U}_t^{(k)}-\boldsymbol{U}_c^{(k)}\right)\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= 1 + C_0 \max_{1\le k\le K}\left[n_c^{-1}\ n_2\ l_k\ O\left(\left(\frac{\log n}{n}\right)^{2/p}\right)\right]^{1/2}\\
&= 1 + C_0 \max_{1\le k\le K}\left[\frac{p}{\log n}\ O\left(\left(\frac{\log n}{n}\right)^{2/p}\right)\right]^{1/2}\\
&= 1 + o(1).
\end{aligned}
$$

Then

$$
\begin{aligned}
&P\left\{\sup_{\boldsymbol{w}\in H_n}\left|\left\|\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}\right\|^2 - \sigma_t^2 tr\big(\boldsymbol{P}_t'(\boldsymbol{w})\boldsymbol{P}_t(\boldsymbol{w})\big)\right| / R_n(\boldsymbol{w}) > \delta\right\}\\
\le\ & P\left\{\sup_{\boldsymbol{w}\in H_n}\sum_{j=1}^K\sum_{k=1}^K w_j w_k\left|\boldsymbol{\zeta}_{\boldsymbol{ta}}'\boldsymbol{P}_t^{(j)\prime}\boldsymbol{P}_t^{(k)}\boldsymbol{\zeta}_{ta} - \sigma_t^2 tr\big(\boldsymbol{P}_t^{(j)\prime}\boldsymbol{P}_t^{(k)}\big)\right| > \delta\xi_n\right\}\\
\le\ & P\left\{\max_{1\le j\le K}\max_{1\le k\le K}\left|\boldsymbol{\zeta}_{ta}'\boldsymbol{P}_t^{(j)\prime}\boldsymbol{P}_t^{(k)}\boldsymbol{\zeta}_{ta} - \sigma_t^2 tr\big(\boldsymbol{P}_t^{(j)\prime}\boldsymbol{P}_t^{(k)}\big)\right| > \delta\xi_n\right\}\\
\le\ & \sum_{j=1}^K\sum_{k=1}^K P\left\{\left|\big\langle\boldsymbol{\zeta}_{ta}, \boldsymbol{P}_t'(\boldsymbol{w}_j^0)\boldsymbol{P}_t(\boldsymbol{w}_k^0)\boldsymbol{\zeta}_{ta}\big\rangle - \sigma_t^2 tr\big(\boldsymbol{P}_t(\boldsymbol{w}_j^0)'\boldsymbol{P}_t(\boldsymbol{w}_k^0)\big)\right| > \delta\xi_n\right\}\\
\le\ & \sum_{j=1}^K\sum_{k=1}^K E\left[\frac{\left\{\big\langle\boldsymbol{\zeta}_{ta}, \boldsymbol{P}_t'(\boldsymbol{w}_j^0)\boldsymbol{P}_t(\boldsymbol{w}_k^0)\boldsymbol{\zeta}_{ta}\big\rangle - \sigma_t^2 tr\big(\boldsymbol{P}_t'(\boldsymbol{w}_j^0)\boldsymbol{P}_t(\boldsymbol{w}_k^0)\big)\right\}^{2G}}{\delta^{2G}\xi_n^{2G}}\right]\\
\le\ & C_7\delta^{-2G}\xi_n^{-2G}\sum_{j=1}^K\sum_{k=1}^K\left[tr\big(\boldsymbol{P}_t'(\boldsymbol{w}_k^0)\boldsymbol{P}_t(\boldsymbol{w}_j^0)\boldsymbol{P}_t'(\boldsymbol{w}_j^0)\boldsymbol{P}_t(\boldsymbol{w}_k^0)\big)\right]^G\\
\le\ & C_7\delta^{-2G}\xi_n^{-2G}\bar{\lambda}^2\big(\boldsymbol{P}_t(\boldsymbol{w}_j^0)\big)\sum_{j=1}^K\sum_{k=1}^K\left[tr\big(\boldsymbol{P}_t'(\boldsymbol{w}_k^0)\boldsymbol{P}_t(\boldsymbol{w}_k^0)\big)\right]^G\\
\le\ & C_7\delta^{-2G}\xi_n^{-2G}K\sum_{k=1}^K\big(R_{n_t}(\boldsymbol{w}_k^0)\big)^G\\
\to\ & 0
\end{aligned}
$$

and

$$
\begin{aligned}
&P\left\{\sup_{\boldsymbol{w}\in H_n}\left|\left\|\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}\right\|^2 - \sigma_c^2 tr\big(\boldsymbol{P}_{\tilde{c}}'(\boldsymbol{w})\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\big)\right| / R_n(\boldsymbol{w}) > \delta\right\}\\
\le\ & P\left\{\sup_{\boldsymbol{w}\in H_n}\sum_{j=1}^K\sum_{k=1}^K w_j w_k\left|\boldsymbol{v}_{ca}'\boldsymbol{P}_{\tilde{c}}^{(j)\prime}\boldsymbol{P}_{\tilde{c}}^{(k)}\boldsymbol{v}_{ca} - \sigma_c^2 tr\big(\boldsymbol{P}_{\tilde{c}}^{(j)\prime}\boldsymbol{P}_{\tilde{c}}^{(k)}\big)\right| > \delta\xi_n\right\}\\
\le\ & P\left\{\max_{1\le j\le K}\max_{1\le k\le K}\left|\boldsymbol{v}_{ca}'\boldsymbol{P}_{\tilde{c}}^{(j)\prime}\boldsymbol{P}_{\tilde{c}}^{(k)}\boldsymbol{v}_{ca} - \sigma_c^2 tr\big(\boldsymbol{P}_{\tilde{c}}^{(j)\prime}\boldsymbol{P}_{\tilde{c}}^{(k)}\big)\right| > \delta\xi_n\right\}\\
\le\ & C_8\delta^{-2G}\xi_n^{-2G}\sum_{j=1}^K\sum_{k=1}^K\left[tr\left\{\boldsymbol{P}_{\tilde{c}}'(\boldsymbol{w}_k^0)\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w}_j^0)\boldsymbol{P}_{\tilde{c}}'(\boldsymbol{w}_j^0)\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w}_k^0)\right\}\right]^G
\end{aligned}
$$

$$\leq C_8 \delta^{-2G} \xi_n^{-2G} \bar{\lambda}\big(\boldsymbol{P}_{\tilde{c}}'(\boldsymbol{w}^*) \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w}^*)\big) \sum_{j=1}^{K} \sum_{k=1}^{K} \left[tr \left\{ \boldsymbol{P}_{\tilde{c}}'(\boldsymbol{w}_k^0) \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w}_k^0) \right\} \right]^G$$

$$\leq C_8' \delta^{-2G} \xi_n^{-2G} K \sum_{k=1}^{K} \big(R_{n_{\tilde{c}}}(\boldsymbol{w}_k^0)\big)^G$$

$$\to 0.$$

Additionally, we can show that

$$P\left\{ \sup_{\boldsymbol{w} \in H_n} |\langle \boldsymbol{A}_t(\boldsymbol{w}), \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w}) \boldsymbol{v}_{ca}) \rangle| / R_n(\boldsymbol{w}) > \delta \right\}$$

$$\leq P\left\{ \sup_{\boldsymbol{w} \in H_n} \left| \boldsymbol{v}_{ca}' \sum_{j=1}^{K} w_j \boldsymbol{P}_{\tilde{c}}^{(j)\prime} \sum_{k=1}^{K} w_k \big(\boldsymbol{f}_t - \boldsymbol{P}_t^{(k)} \boldsymbol{f}_{ta}\big) \right| > \delta \xi_n \right\}$$

$$\leq P\left\{ \sum_{j=1}^{K} \sum_{k=1}^{K} w_j w_k \left| \boldsymbol{v}_{ca}' \boldsymbol{P}_{\tilde{c}}^{(j)\prime} \big(\boldsymbol{f}_t - \boldsymbol{P}_t^{(k)} \boldsymbol{f}_{ta}\big) \right| > \delta \xi_n \right\}$$

$$\leq P\left\{ \max_{1 \leq j \leq K} \max_{1 \leq k \leq K} \left| \boldsymbol{v}_{ca}' \boldsymbol{P}_{\tilde{c}}^{(j)\prime} \big(\boldsymbol{f}_t - \boldsymbol{P}_t^{(k)} \boldsymbol{f}_{ta}\big) \right| > \delta \xi_n \right\}$$

$$\leq \sum_{j=1}^{K} \sum_{k=1}^{K} P\left\{ \left| \big\langle \boldsymbol{A}_t(\boldsymbol{w}_j^0), \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w}_k^0) \boldsymbol{v}_{ca}) \big\rangle \right| > \delta \xi_n \right\}$$

$$\leq \sum_{j=1}^{K} \sum_{k=1}^{K} E\left[\frac{\big\langle \boldsymbol{A}_t(\boldsymbol{w}_j^0), \boldsymbol{P}_{\tilde{c}}(\boldsymbol{w}_k^0) \boldsymbol{v}_{ca}) \big\rangle^{2G}}{\delta^{2G} \xi_n^{2G}} \right]$$

$$\leq C_9 \delta^{-2G} \xi_n^{-2G} \sum_{j=1}^{K} \sum_{k=1}^{K} \left\| \boldsymbol{P}_{\tilde{c}}^{(j)\prime} \big(\boldsymbol{f}_t - \boldsymbol{P}_t^{(k)} \boldsymbol{f}_{ta}\big) \right\|^{2G}$$

$$\leq C_9 \delta^{-2G} \xi_n^{-2G} \sum_{j=1}^{K} \sum_{k=1}^{K} \bar{\lambda}\big(\boldsymbol{P}_{\tilde{c}}^{(j)\prime} \boldsymbol{P}_{\tilde{c}}^{(j)}\big) \left\| \boldsymbol{A}_t(\boldsymbol{w}_k^0) \right\|^{2G}$$

$$\leq C_9' \delta^{-2G} \xi_n^{-2G} K \sum_{k=1}^{K} \big(R_{n_{\tilde{c}}}(\boldsymbol{w}_k^0)\big)^G \to 0.$$

Similarly, we have

$$P\left\{ \sup_{\boldsymbol{w} \in H_n} \left| \big\langle \boldsymbol{A}_t(\boldsymbol{w}), \boldsymbol{P}_t(\boldsymbol{w}) \boldsymbol{\zeta}_{ta} \big\rangle \right| / R_n(\boldsymbol{w}) > \delta \right\}$$

$$\leq C_{10} \delta^{-2G} \xi_n^{-2G} \sum_{j=1}^{K} \sum_{k=1}^{K} \bar{\lambda}\big(\boldsymbol{P}_t^{(j)\prime} \boldsymbol{P}_t^{(j)}\big) \left\| \boldsymbol{A}_t(\boldsymbol{w}_k^0) \boldsymbol{f}_t \right\|^{2G}$$

$$\leq C_{10}' \delta^{-2G} \xi_n^{-2G} K \sum_{k=1}^{K} \big(R_{n_t}(\boldsymbol{w}_k^0)\big)^G \to 0,$$

$$
\begin{aligned}
&P\left\{\sup_{\boldsymbol{w}\in H_n}\left|\left\langle \boldsymbol{A}_{\tilde{c}}(\boldsymbol{w}),\boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta}\right\rangle\right|/R_n(\boldsymbol{w})>\delta\right\}\\
&\le C_{11}\delta^{-2G}\xi_n^{-2G}\sum_{j=1}^K\sum_{k=1}^K\bar{\lambda}\big(\boldsymbol{P}_t^{(j)'}\boldsymbol{P}_t^{(j)}\big)\left\|\boldsymbol{A}_{\tilde{c}}(\boldsymbol{w}_k^0)\right\|^{2G}\\
&\le C_{11}'\delta^{-2G}\xi_n^{-2G}K\sum_{k=1}^K\big(R_{\tilde{c}}(\boldsymbol{w}_k^0)\big)^G\to 0,
\end{aligned}
$$

$$
\begin{aligned}
&P\left\{\sup_{\boldsymbol{w}\in H_n}\left|\left\langle \boldsymbol{A}_{\tilde{c}}(\boldsymbol{w}),\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}\right\rangle\right|/R_n(\boldsymbol{w})>\delta\right\}\\
&\le C_{12}\delta^{-2G}\xi_n^{-2G}\sum_{j=1}^K\sum_{k=1}^K\bar{\lambda}\big(\boldsymbol{P}_{\tilde{c}}^{(j)'}\boldsymbol{P}_{\tilde{c}}^{(j)}\big)\left\|\boldsymbol{A}_{\tilde{c}}(\boldsymbol{w}_k^0)\right\|^{2G}\\
&\le C_{12}'\delta^{-2G}\xi_n^{-2G}K\sum_{k=1}^K\big(R_{n_{\tilde{c}}}(\boldsymbol{w}_k^0)\big)^G\to 0
\end{aligned}
$$

and

$$
\begin{aligned}
&P\left\{\sup_{\boldsymbol{w}\in H_n}\left|\left\langle \boldsymbol{P}_t(\boldsymbol{w})\boldsymbol{\zeta}_{ta},\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w})\boldsymbol{v}_{ca}\right\rangle\right|/R_n(\boldsymbol{w})>\delta\right\}\\
&\le P\left\{\sup_{\boldsymbol{w}\in H_n}\sum_{j=1}^K\sum_{k=1}^K w_jw_k\left|\boldsymbol{\zeta}_{ta}'\boldsymbol{P}_t^{(j)'}\boldsymbol{P}_{\tilde{c}}^{(k)}\boldsymbol{v}_{ca}\right|>\delta\xi_n\right\}\\
&\le P\left\{\max_{1\le j\le K}\max_{1\le k\le K}\left|\boldsymbol{\zeta}_{ta}'\boldsymbol{P}_t^{(j)'}\boldsymbol{P}_{\tilde{c}}^{(k)}\boldsymbol{v}_{ca}\right|>\delta\xi_n\right\}\\
&\le \sum_{j=1}^K\sum_{k=1}^K P\left\{\left|\boldsymbol{\zeta}_{ta}'\boldsymbol{P}_t^{(j)'}\boldsymbol{P}_{\tilde{c}}^{(k)}\boldsymbol{v}_{ca}\right|>\delta\xi_n\right\}\\
&\le \sum_{j=1}^K\sum_{k=1}^K\frac{E\left|\boldsymbol{\zeta}_{ta}'\boldsymbol{P}_t^{(j)'}\boldsymbol{P}_{\tilde{c}}^{(k)}\boldsymbol{v}_{ca}\right|^2}{\delta^2\xi_n^2}\\
&\le \sum_{j=1}^K\sum_{k=1}^K\frac{E\left\|\boldsymbol{P}_t^{(j)}\boldsymbol{\zeta}_{ta}\right\|^2E\left\|\boldsymbol{P}_{\tilde{c}}^{(k)}\boldsymbol{v}_{ca}\right\|^2}{\delta^2\xi_n^2}\\
&\le C_{13}\delta^{-2}\xi_n^{-2}\sum_{j=1}^K\sum_{k=1}^K tr\left|\boldsymbol{P}_t(\boldsymbol{w}_j^0)\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w}_k^0)\right|^2\\
&\le C_{13}\delta^{-2}\xi_n^{-2}\bar{\lambda}(\boldsymbol{P}_t^2(\boldsymbol{w}^*))\sum_{j=1}^K\sum_{k=1}^K\left[tr\big(\boldsymbol{P}_{\tilde{c}}'(\boldsymbol{w}_j^0)\boldsymbol{P}_{\tilde{c}}(\boldsymbol{w}_j^0)\big)\right]\\
&\le C_{13}'\delta^{-2}\xi_n^{-2}K\sum_{k=1}^K\big(R_{n_{\tilde{c}}}(\boldsymbol{w}_k^0)\big)\to 0.
\end{aligned}
$$

By the above results, it is readily seen that the right hand side of (A.26) converges to zero in probability. From (A.26) and noting that

$$\begin{aligned}
&\sup_{\boldsymbol{w}\in H_n}\left|\tilde{B}(\boldsymbol{w})\right|/R_n(\boldsymbol{w})\\
&=\sup_{\boldsymbol{w}\in H_n}\left|\underline{B}(\boldsymbol{w})-L_n(\boldsymbol{w})\right|/R_n(\boldsymbol{w})\\
&=\sup_{\boldsymbol{w}\in H_n}\left|\left\|\boldsymbol{\Delta}+\boldsymbol{b}-\widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2-\left\|\boldsymbol{\Delta}-\widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2\right|/R_n(\boldsymbol{w})\\
&\le\sup_{\boldsymbol{w}\in H_n}\left[2(\underline{B}(\boldsymbol{w})\left\|\boldsymbol{b}\right\|^2)^{1/2}\right]/R_n(\boldsymbol{w})+\sup_{\boldsymbol{w}\in H_n}\left\|\boldsymbol{b}\right\|^2/R_n(\boldsymbol{w})\to_p 0, \quad \text{(A.27)}
\end{aligned}$$

we can see that

$$\begin{aligned}
&\sup_{\boldsymbol{w}\in H_n}\left|L_n(\boldsymbol{w})/R_n(\boldsymbol{w})-1\right|\\
&=\sup_{\boldsymbol{w}\in H_n}\left|[\{\underline{B}(\boldsymbol{w})-L_n(\boldsymbol{w})\}-\{\underline{B}(\boldsymbol{w})-R_n(\boldsymbol{w})\}]\right|/R_n(\boldsymbol{w})\\
&=\sup_{\boldsymbol{w}\in H_n}\left|\underline{B}(\boldsymbol{w})-L_n(\boldsymbol{w})\right|/R_n(\boldsymbol{w})+\sup_{\boldsymbol{w}\in H_n}\left|\underline{B}(\boldsymbol{w})-R_n(\boldsymbol{w})\right|/R_n(\boldsymbol{w})\to_p 0,
\end{aligned}$$

which is (A.13). Thus, (A.3), (A.12) and (A.13) are proved, and this completes the proof of Theorem 2. □

A.3 Proof of Theorem 3

Let

$$\begin{aligned}
\widehat{C}_n(\boldsymbol{w})&=\left\|\tilde{\boldsymbol{Y}}-\widehat{\boldsymbol{\Delta}}(\boldsymbol{w})\right\|^2+2tr\big(\boldsymbol{P}_{t0}(\boldsymbol{w})\big)\widehat{\sigma}^2_{K_t^*}+2tr\big(\boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\big)\widehat{\sigma}^2_{K_c^*}\\
&=C_n(\boldsymbol{w})+2tr\big(\boldsymbol{P}_{t0}(\boldsymbol{w})\big)(\widehat{\sigma}^2_{K_t^*}-\sigma_t^2)+2tr\big(\boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\big)(\widehat{\sigma}^2_{K_c^*}-\sigma_c^2). \quad \text{(A.28)}
\end{aligned}$$

By (A.28) and Theorem 2, Theorem 3 is valid if the following conditions hold

$$\sup_{\boldsymbol{w}\in H_n} tr\big(\boldsymbol{P}_{t0}(\boldsymbol{w})\big)\left|\widehat{\sigma}^2_{K_t^*}-\sigma_t^2\right|/R_n(\boldsymbol{w})\to_p 0 \quad \text{(A.29)}$$

and

$$\sup_{\boldsymbol{w}\in H_n} tr\big(\boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\big)\left|\widehat{\sigma}^2_{K_c^*}-\sigma_c^2\right|/R_n(\boldsymbol{w})\to_p 0. \quad \text{(A.30)}$$

We need only to verify (A.30), (A.29) can be proved in the same way.

It should be mentioned that our assumption of $\boldsymbol{U}$ being non-stochastic instead of allowing it to be stochastic has no impact on our proof, because all of our technical conditions related to $\boldsymbol{U}$ hold almost surely. Let $r_{\tilde{c}}=tr\big(\boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\big)$. Then we have

$$r_{\tilde{c}}\le\bar{\lambda}\big(\boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\big)\ l_{K^*}=\big(1+o_p(1)\big)l_{K^*} \quad \text{(A.31)}$$

and

$$\begin{aligned}
&\sup_{\boldsymbol{w}\in H_n} tr\big(\boldsymbol{P}_{\tilde{c}0}(\boldsymbol{w})\big)\left|\widehat{\sigma}^2_{K_c^*}-\sigma_c^2\right|/R_n(\boldsymbol{w})\\
&\leq \frac{r_{\tilde{c}}}{\xi_n}\left|\widehat{\sigma}^2_{K_c^*}-\sigma_c^2\right| = \frac{r_{\tilde{c}}}{\xi_n}\left|\frac{\boldsymbol{Y}_{ca}'\big(\boldsymbol{I}-\boldsymbol{P}_{ca}^{(K^*)}\big)\boldsymbol{Y}_{ca}}{n_c-l_{K^*}}-\sigma_c^2\right|\\
&\leq \frac{r_{\tilde{c}}}{n_2-l_{K^*}}\frac{\boldsymbol{f}_{ca}'\big(\boldsymbol{I}-\boldsymbol{P}_{ca}^{(K^*)}\big)\boldsymbol{f}_{ca}}{\xi_n}+\frac{2r_{\tilde{c}}\left|\boldsymbol{f}_{ca}'\big(\boldsymbol{I}-\boldsymbol{P}_{ca}^{(K^*)}\big)\boldsymbol{v}_{ca}\right|}{\xi_n(n_c-l_{K^*})}\\
&\quad+\frac{r_{\tilde{c}}\left|\boldsymbol{v}_{ca}'\big(\boldsymbol{I}-\boldsymbol{P}_{ca}^{(K^*)}\big)\boldsymbol{v}_{ca}-(n_c-l_{K^*})\sigma_c^2\right|}{\xi_n(n_c-l_{K^*})}. \qquad (A.32)
\end{aligned}$$

It suffices to show that the three terms on the right hand of (A.32) tend to zero in probability. Using Condition (25), we obtain

$$\frac{\boldsymbol{f}_{ca}'\big(\boldsymbol{I}-\boldsymbol{P}_{ca}^{(K^*)}\big)\boldsymbol{f}_{ca}}{\xi_n^2}\to 0. \qquad (A.33)$$

Then from (24), (26), (A.31) and (A.33), it is seen that

$$\begin{aligned}
&\frac{r_{\tilde{c}}}{n_c-l_{K^*}}\frac{\boldsymbol{f}_{ca}'\big(\boldsymbol{I}-\boldsymbol{P}_{ca}^{(K^*)}\big)\boldsymbol{f}_{ca}}{\xi_n}\\
&\leq\left[\frac{r_{\tilde{c}}^2}{n_c-l_{K^*}}\frac{\boldsymbol{f}_{ca}'\big(\boldsymbol{I}-\boldsymbol{P}_{ca}^{(K^*)}\big)\boldsymbol{f}_{ca}}{\xi_n^2}\frac{\boldsymbol{f}_{ca}'\boldsymbol{f}_{ca}}{n_c-l_{K^*}}\right]^{1/2}\to 0 \qquad (A.34)
\end{aligned}$$

and

$$\begin{aligned}
&P\left\{\frac{2r_{\tilde{c}}\left|\boldsymbol{f}_{ca}'(\boldsymbol{I}-\boldsymbol{P}_{ca}^{(K^*)})\boldsymbol{v}_{ca}\right|}{\xi_n(n_c-l_{K^*})}>\delta\right\}\\
&\leq E\left[\boldsymbol{f}_{ca}'\big(\boldsymbol{I}-\boldsymbol{P}_{ca}^{(K^*)}\big)\boldsymbol{v}_{ca}\right]^2\frac{4r_{\tilde{c}}^2}{\delta^2\xi_n^2(n_c-l_{K^*})^2}\\
&\leq\frac{C_{14}r_{\tilde{c}}^2\boldsymbol{f}_{ca}'\big(\boldsymbol{I}-\boldsymbol{P}_{ca}^{(K^*)}\big)\boldsymbol{f}_{ca}}{\delta^2\xi_n^2(n_c-l_{K^*})^2}\to 0. \qquad (A.35)
\end{aligned}$$

By (A.35), it can be seen the second term on the right hand side of (A.32) tends to zero in probability. From

$$\begin{aligned}
&P\left\{\frac{r_{\tilde{c}}|\boldsymbol{v}_{ca}'\big(\boldsymbol{I}-\boldsymbol{P}_{ca}^{(K^*)}\big)\boldsymbol{v}_{ca}-(n_c-l_{K^*})\sigma_c^2|}{\xi_n(n_c-l_{K^*})}>\delta\right\}\\
&\leq E\left[\boldsymbol{v}_{ca}'\big(\boldsymbol{I}-\boldsymbol{P}_{ca}^{(K^*)}\big)\boldsymbol{v}_{ca}-(n_c-l_{K^*})\sigma_c^2\right]^2\frac{r_{\tilde{c}}^2}{\delta^2\xi_n^2(n_c-l_{K^*})^2}\\
&\leq\frac{C_{15}r_{\tilde{c}}^2(n_c-l_tK^*)}{\delta^2\xi_n^2(n_c-l_{K^*})^2}\to 0, \qquad (A.36)
\end{aligned}$$

we see the third term on the right hand side of (A.32) tends to zero in probability. Combining (A.32)–(A.36), it is straightforward to see that (A.30) holds. This completes the proof of Theorem 3. □

A.4 An example in which Condition (14) is satisfied

Consider the data-generating process

$$Y_i = [f_t(z_i) + \zeta_i]\, I(T_i = t) + [f_c(z_i) + v_i]\, I(T_i = c), \qquad 1 \le i \le n,$$

where $f_t(z_i) = -f_c(z_i) = \sum_{s=1}^{\infty} \theta_s \cos((s-1)z_i)/s$, $z_i = 2\pi(i-1)/n$, and ζ_i and v_i are i.i.d. $N(0, \sigma^2)$. Suppose that $\theta_s = s^{-7/12}$, and we smooth across estimators from models that comprise subsets of the first p ($p < S$) regressors, where S is the largest integer that is no greater than $n/2$. Altogether there are $K = 2^{p-1}$ approximating models if we assume all models contain the intercept.

From the proof of Example 1 in Wan et al. (2010), it can be seen that $\frac{np^{-13/6}}{\xi_n} = o(1)$. Hence Condition (14) holds if $p = O\left((\log n)^{1/5}\right)$.

A.5 Results of Design 2

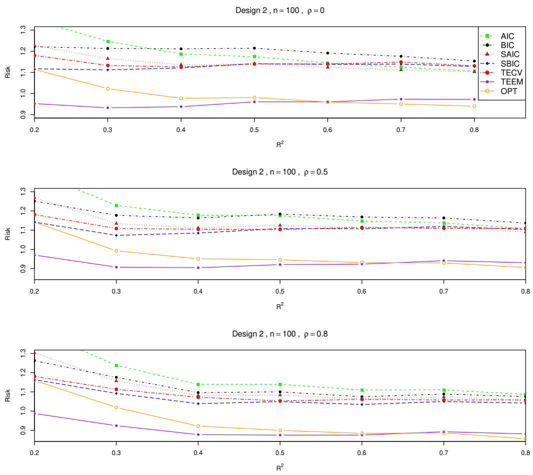

Fig. 6: Results of simulations under Design 2 with $n = 100$.

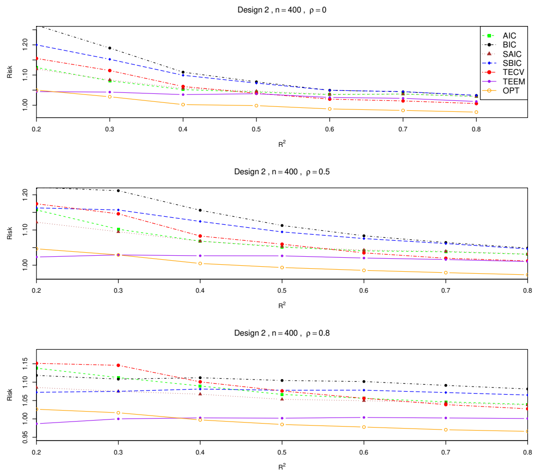

Fig. 7: Results of simulations under Design 2 with $n = 400$.

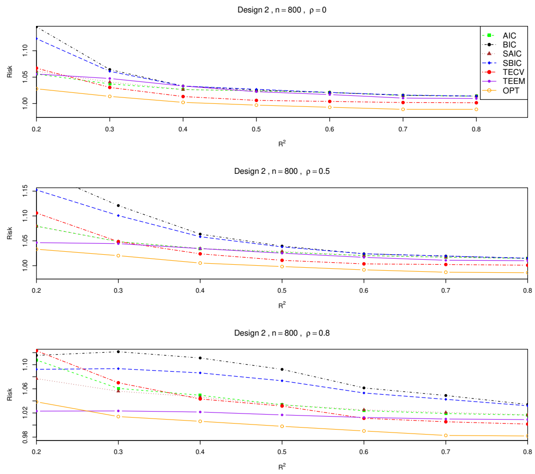

Fig. 8: Results of simulations under Design 2 with $n = 800$.

References

Wan, A. T., Zhang, X., Zou, G. (2010). Least squares model averaging by Mallows criterion. *Journal of Econometrics*, *156*(2), 277–283.

Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory of Probability and Its Applications*, *5*(3), 302–305.

Zhang, X., Wan, A. T., Zou, G. (2013). Model averaging by jackknife criterion in models with dependent data. *Journal of Econometrics*, *174*(2), 82–94.