Appendix

A.1 Proof of Theorem 1

Let \( \mathbf{b} = (b_1, \ldots, b_n)' \) and \( \mathbf{\eta} = (\eta_1, \ldots, \eta_n)' \). Then

\[
C_n(w) = \left\| \mathbf{Y} - \hat{\Delta}(w) \right\|^2 + 2\sigma_t^2 \text{tr}(P_0(w)) + 2\sigma_c^2 \text{tr}(P_\Theta(w))
= \left\| \Delta + \mathbf{b} + \mathbf{\eta} - \hat{\Delta}(w) \right\|^2 + 2\sigma_t^2 \text{tr}(P_0(w)) + 2\sigma_c^2 \text{tr}(P_\Theta(w))
= 2 \left[ \langle \Delta + \mathbf{b} - \hat{\Delta}(w), \mathbf{\eta} \rangle + \sigma_t^2 \text{tr}(P_0(w)) + \sigma_c^2 \text{tr}(P_\Theta(w)) \right]
+ \left\| \Delta + \mathbf{b} - \hat{\Delta}(w) \right\|^2 + \| \mathbf{\eta} \|^2
\]
\[
= \bar{A}(w) + \bar{B}(w) + \| \mathbf{\eta} \|^2. \tag{A.1}
\]
Let us first consider the second term on the right-hand-side of (A.1)

\[ B(w) = \|\Delta + b - \hat{\Delta}(w)\|^2 = \|b\|^2 + 2\langle \Delta - \hat{\Delta}(w), b \rangle + \|\Delta - \hat{\Delta}(w)\|^2. \]

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

\[
E \|b\|^2|U \leq E \|\Delta - \hat{\Delta}(w)\|^2|U \leq \frac{E \|b\|^2|U}{\xi_n} = \frac{n}{\xi_n} O\left(\frac{p \log n}{n^{1/p}}\right)^{1/(2p)} \xi_n^{-1} = o(1), \text{ a.s.} \tag{A.2}
\]

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

\[
E \langle \Delta - \hat{\Delta}(w), b \rangle |U \leq E \|\Delta - \hat{\Delta}(w)\|^2 |U\right) \]

\[
= E \left[ \left( \frac{n}{\log n} \right)^{1/(2p)} \right] |U\right)
\]

\[
= O \left( \frac{p^2 (\log n)^{1/(2p)}}{\xi_n^{1/(2p)}} \right) \xi_n^{-1}
\]

\[
= o(1), \text{ a.s.} \tag{A.3}
\]

Then

\[
E(B(w)|U) = E \left[ \left( \frac{n}{\log n} \right)^{1/(2p)} \right] + 2E \langle \Delta - \hat{\Delta}(w), b \rangle + \|b\|^2 |U\right)
\]

\[
\leq R_n(w) + E \left( \left( \frac{n}{\log n} \right)^{1/(2p)} \right) \left\{ E \left( \frac{\|b\|^2 |U\right)|U\right) \right\}^{1/2}
\]
This completes the proof of Theorem 1.

Let \( \mathbf{c}_{1a} = Y_{ta} - f_{1a}, \mathbf{v}_{1a} = Y_{ta} - f_{ca}, \hat{f}^{(k)}_t = P^{(k)}_t Y_{ta} \) and \( \hat{f}^{(k)}_c = P^{(k)}_c Y_{ca} \).

From (12), we obtain the model average estimator

\[
\hat{\boldsymbol{\Delta}}(\mathbf{w}) = \sum_{k=1}^{K} w_k \left( \hat{f}^{(k)}_t - \hat{f}^{(k)}_c \right)
\]

which yields

\[
\Delta + b - \hat{\boldsymbol{\Delta}}(\mathbf{w}) = f_t - f_c - \hat{\boldsymbol{\Delta}}(\mathbf{w})
= (f_t - P_t(\mathbf{w})f_{1a}) - (f_c - P_c(\mathbf{w})f_{ca}) - P_t(\mathbf{w})\zeta_{ta} + P_c(\mathbf{w})\mathbf{v}_{ca}
:= A_t(\mathbf{w}) - A_c(\mathbf{w}) - P_t(\mathbf{w})\zeta_{ta} + P_c(\mathbf{w})\mathbf{v}_{ca}.
\]

Let \( \mathbf{c}_t = (\mathbf{c}_1, \ldots, \mathbf{c}_n) \) and \( \mathbf{v}_c = (\mathbf{v}_1, \ldots, \mathbf{v}_n) \). From (5) and (6), we have

\[
E(\mathbf{c}_t^2|\mathbf{U}) = \sigma_t^2 + \sigma_c^2 \quad \text{and} \quad E(\mathbf{c}_t^2|\mathbf{U}) = 0.
\]

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

\[
E(\mathbf{A}(\mathbf{w})|\mathbf{U}) = 2E\left( \left\langle A_t(\mathbf{w}) - A_c(\mathbf{w}), \zeta_{ta} + P_t(\mathbf{w})\mathbf{v}_{ca}, \mathbf{c}_t - \mathbf{c}_c \right\rangle |\mathbf{U} \right)
+ \sigma_t^2 tr(P_{10}(\mathbf{w})) + \sigma_c^2 tr(P_{20}(\mathbf{w}))
= 2E\left( \left\langle A_t(\mathbf{w}) - A_c(\mathbf{w}), \eta \right\rangle |\mathbf{U} \right)
+ 2E\left( \left\langle -P_t(\mathbf{w})\zeta_{ta}, \mathbf{c}_t \right\rangle |\mathbf{U} \right)
+ 2\sigma_t^2 tr(P_{10}(\mathbf{w})) + 2E\left( \left\langle P_t(\mathbf{w})\mathbf{v}_{ca}, -\mathbf{v}_c \right\rangle |\mathbf{U} \right)
+ 2\sigma_c^2 tr(P_c(\mathbf{w}))
\]

and

\[
E(\|\boldsymbol{\eta}\|^2|\mathbf{U}) = n_2(\sigma_t^2 + \sigma_c^2).
\]

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

\[
E(C_{\mathbf{w}}(\mathbf{w})|\mathbf{U}) = E(\mathbf{A}(\mathbf{w})|\mathbf{U}) + E(\mathbf{B}(\mathbf{w})|\mathbf{U}) + E(\|\boldsymbol{\eta}\|^2|\mathbf{U})
= R_n(\mathbf{w}) (1 + o(1)) + n_2(\sigma_t^2 + \sigma_c^2), \quad \text{a.s.}
\]
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 \( U \) is non-stochastic. The proof that follows also applies to the case of stochastic \( U \) because all of the technical conditions imposed on \( U \) hold almost surely. In the following proof, we assume that \( C_1, \ldots, C_{15}, C'_8, \ldots, C'_{13} \) are distinct constants.

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

\[
C_n(w) = L_n(w) + A(w) + (B(w) - L_n(w)) + \|\eta\|^2 =: L_n(w) + A(w) + \tilde{B}(w) + \|\eta\|^2. \tag{A.10}
\]

Theorem 2 is valid if the following hold

\[
sup_{w \in H_n} \left| \frac{A(w)}{R_n(w)} \right| \to_p 0, \tag{A.11}
\]

\[
sup_{w \in H_n} \left| \frac{\tilde{B}(w)}{R_n(w)} \right| \to_p 0 \tag{A.12}
\]

and

\[
sup_{w \in H_n} \left| \frac{L_n(w)}{R_n(w)} - 1 \right| \to_p 0. \tag{A.13}
\]

From (A.5), we have

\[
A(w) = 2 \langle \Delta + b - \tilde{A}(w), \eta \rangle + 2\sigma_t^2 \text{tr} \left( P_t(w) \right) = 2 \langle f_t - f_c - P_t(w) f_{ta} + P_c(w) f_{ca} - P_t(w) \xi_{ta} + P_c(w) v_{ca}, \eta \rangle + 2\sigma_t^2 \text{tr} \left( P_{t0}(w) \right) = 2 \langle f_t - P_t(w) f_{ta}, \eta \rangle - 2 \langle f_c - P_c(w) f_{ca}, \eta \rangle - 2 \left\{ \langle P_t(w) \xi_{ta}, \eta \rangle - \sigma_t^2 \text{tr} \left( P_{t0}(w) \right) \right\} + 2 \left\{ \langle P_c(w) v_{ca}, \eta \rangle + \sigma_c^2 \text{tr} \left( P_{c0}(w) \right) \right\} = 2 \langle A_t(w), \eta \rangle - 2 \langle A_c(w), \eta \rangle - 2 \left\{ \langle P_t(w) \xi_{ta}, \eta \rangle - \sigma_t^2 \text{tr} \left( P_{t0}(w) \right) \right\} + 2 \left\{ \langle P_c(w) v_{ca}, \eta \rangle + \sigma_c^2 \text{tr} \left( P_{c0}(w) \right) \right\}.
\]

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

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

\[
sup_{w \in H_n} \left| \langle A_c(w), \eta \rangle \right| / R_n(w) \to_p 0, \tag{A.16}
\]

\[
sup_{w \in H_n} \left| \langle P_t(w) \xi_{ta}, \eta \rangle - \sigma_t^2 \text{tr} \left( P_{t0}(w) \right) \right| / R_n(w) \to_p 0. \tag{A.17}
\]
and
\[ \sup_{w \in H_n} \left| \langle P_c(w) v_c, \eta \rangle + \sigma^2_c \text{tr} \left( P_{c0}(w) \right) \right| / R_n(w) \to_p 0. \]  
(A.18)

Note that Condition (17) implies that
\[ E \left( (\zeta_m^t)^4 G | U \right) \leq \kappa < \infty \text{ and } E \left( (v_c^t)^4 G | U \right) \leq \kappa < \infty, m = 1, \ldots, n_2. \]  
(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 $\delta > 0$,
\[
P \left\{ \sup_{w \in H_n} \left| \langle A_t(w), \eta \rangle \right| / R_n(w) > \delta \right\}
\leq P \left\{ \sup_{w \in H_n} \sum_{k=1}^K w_k \left| \eta' \left( f_t - P_t^{(k)} f_{ta} \right) \right| > \delta \xi_n \right\}
\leq P \left\{ \max_{1 \leq k \leq K} \left| \eta' \left( f_t - P_t^{(k)} f_{ta} \right) \right| > \delta \xi_n \right\}
= P \left\{ \left\{ \left| \langle \eta, A_t(w_1^0) \rangle \right| > \delta \xi_n \right\} \cup \left\{ \left| \langle \eta, A_t(w_2^0) \rangle \right| > \delta \xi_n \right\} \cup \cdots \cup \left\{ \left| \langle \eta, A_t(w_2^0) \rangle \right| > \delta \xi_n \right\} \right\}
\leq \sum_{k=1}^K P \left\{ \left| \langle \eta, A_t(w_k^0) \rangle \right| > \delta \xi_n \right\}
\leq \sum_{k=1}^K E \left( \left| \langle \eta, A_t(w_k^0) \rangle \right|^2 G \right) / \delta^2 \xi_n^2 G
\leq C_1 \delta^{-2G} \xi_n^{-2G} \sum_{k=1}^K \left| A_t(w_k^0) \right|^2 G \rightarrow 0.

Similarly, for (A.16), we obtain
\[
P \left\{ \sup_{w \in H_n} \left| \langle A_c(w), \eta \rangle \right| / R_n(w) > \delta \right\} \leq C_2 \delta^{-2G} \xi_n^{-2G} \sum_{k=1}^K \left| A_c(w_k^0) \right|^2 G \rightarrow 0.

Now, let us prove (A.17). As
\[
P \left\{ \sup_{w \in H_n} \left| \langle P_t(w) \zeta_{ta}, \eta \rangle - \sigma^2_c \text{tr} \left( P_{t0}(w) \right) \right| / R_n(w) > \delta \right\}
= P \left\{ \sup_{w \in H_n} \left| \langle P_t(w) \zeta_{ta}, \zeta_t - v_c \rangle - \sigma^2_c \text{tr} \left( P_{t0}(w) \right) \right| / R_n(w) > \delta \right\}
\leq P \left\{ \sup_{w \in H_n} \left| \langle P_t(w) \zeta_{ta}, \zeta_t \rangle - \sigma^2_c \text{tr} \left( P_{t0}(w) \right) \right| / R_n(w) > \delta \right\}
+ P \left\{ \sup_{w \in H_n} \left| \langle P_t(w) \zeta_{ta}, v_c \rangle \right| / R_n(w) > \delta \right\}. \]  
(A.20)
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

\[
P \left\{ \sup_{w \in H_n} \left| \langle P_t(w)\zeta_{ta}, \zeta_t \rangle - \sigma_t^2 tr(P_t(w)) \right| / R_n(w) > \delta \right\}
\]

\[
\leq \sum_{k=1}^{K} P \left\{ \left| \langle P_t(w_k^0)\zeta_{ta}, \zeta_t \rangle - \sigma_t^2 tr(P_t(w_k^0)) \right| > \delta \xi_n \right\}
\]

\[
\leq \sum_{k=1}^{K} P \left\{ \left| \langle P_{t0}(w_k^0)\zeta_{ta}, \zeta_{ta} \rangle - \sigma_t^2 tr(P_{t0}(w_k^0)) \right| > \delta \xi_n \right\}
\]

\[
\leq \sum_{k=1}^{K} E \left[ \frac{\left\{ \langle P_{t0}(w_k^0)\zeta_{ta}, \zeta_{ta} \rangle - \sigma_t^2 tr(P_{t0}(w_k^0)) \right\}^2}{\delta^2 \xi_n^2} \right]
\]

\[
\leq C_3 \delta^{-2G} \xi_n^{-2G} \sum_{k=1}^{K} \left[ tr(P_{t0}(w_k^0)P_{t0}(w_k^0)) \right]^G
\]

\[
= C_3 \delta^{-2G} \xi_n^{-2G} \sum_{k=1}^{K} \left[ tr(P_{t}(w_k^0)P_{t}(w_k^0)) \right]^G
\]

\[
\rightarrow 0.
\]

while the second term is given by

\[
P \left\{ \sup_{w \in H_n} \left| \langle P_t(w)\zeta_{ta}, v_c \rangle \right| / R_n(w) > \delta \right\}
\]

\[
\leq \sum_{k=1}^{K} P \left\{ \left| \langle P_t(w_k^0)\zeta_{ta}, v_c \rangle - E\langle P_t(w_k^0)\zeta_{ta}, v_c \rangle \right| > \delta \xi_n \right\}
\]

\[
= \sum_{k=1}^{K} P \left[ \left| \langle P_t(w_k^0)\zeta_{ta}, v_c \rangle \right| > \delta \xi_n \right]
\]

\[
\leq \sum_{k=1}^{K} E \left[ \frac{\langle P_t(w_k^0)\zeta_{ta}, v_c \rangle^2}{\delta^2 \xi_n^2} \right]
\]

\[
\leq \sum_{k=1}^{K} E \left[ \frac{\langle \zeta_{ta}P_t(w_k^0)v_c, v_cP_t(w_k^0)\zeta_{ta} \rangle}{\delta^2 \xi_n^2} \right]
\]

\[
\leq C_4 \delta^{-2G} \xi_n^{-2G} \sum_{k=1}^{K} \left[ tr(P_t(w_k^0)P_t(w_k^0)) \right]
\]

\[
\rightarrow 0.
\]

(A.21)

Hence (A.17) is valid. Similarly, for (A.18), it is observed that

\[
P \left\{ \sup_{w \in H_n} \left| \langle P_c(w)v_{ca}, \eta \rangle + \sigma^2 tr(P_c(w)) \right| / R_n(w) > \delta \right\}
\]
Given (A.15)–(A.18), it is straightforward to see that (A.3) is valid. Now, to prove (A.12), let us first consider

$$B(w) = \left\| \Delta + b - \hat{\Delta}(w) \right\|_2$$

$$= \left\| A_t(w) - A_\hat{z}(w) \right\|^2 + \left\| P_t(w)\zeta_{ta} - P_\hat{c}(w)v_{ca} \right\|^2$$

$$-2\left(A_t(w) - A_\hat{z}(w), P_t(w)\zeta_{ta} - P_\hat{c}(w)v_{ca} \right)$$

Define

$$\hat{R}_n(w) := E(B(w)|U) = E \left[ \left\| A_t(w) - A_\hat{z}(w) \right\|^2 + \left\| P_t(w)\zeta_{ta} - P_\hat{c}(w)v_{ca} \right\|^2$$

$$-2\left(A_t(w) - A_\hat{z}(w), P_t(w)\zeta_{ta} - P_\hat{c}(w)v_{ca} \right) | U \right]$$

$$= \left\| A_t(w) - A_\hat{z}(w) \right\|^2 + \sigma_t^2 tr\left(P_t(w)P_t(w)\right)$$

$$+ \sigma_\hat{c}^2 tr\left(P_\hat{c}(w)P_\hat{c}(w)\right).$$

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

$$\sup_{w \in H_n} \frac{\hat{R}_n(w) - R_n(w)}{R_n(w)}$$

$$= \sup_{w \in H_n} E \left[ \left( \left\| \Delta + b - \hat{\Delta}(w) \right\|^2 - \left\| \Delta - \hat{\Delta}(w) \right\|^2 \right) | U \right] / R_n(w)$$

$$= \sup_{w \in H_n} E \left[ 2 \left( \left\| \Delta - \hat{\Delta}(w), b \right\|^2 \right) | U \right] / R_n(w)$$

$$= o(1).$$

(A.25)

By (A.23)–(A.25), we have

$$\sup_{w \in H_n} \left| \frac{B(w)}{R_n(w)} - 1 \right|$$

$$= \sup_{w \in H_n} \left\{ \left| \frac{B(w) - \hat{R}_n(w)}{R_n(w)} \right| + \left| \frac{\hat{R}_n(w) - R_n(w)}{R_n(w)} \right| \right\} / R_n(w)$$

$$= \sup_{w \in H_n} \left\{ \left\| P_t(w)\zeta_{ta} - P_\hat{c}(w)v_{ca} \right\|^2 - \sigma_t^2 tr\left(P_t(w)P_t(w)\right) - \sigma_\hat{c}^2 tr\left(P_\hat{c}(w)P_\hat{c}(w)\right) \right\} / R_n(w)$$
Recognising that \( A_t(w) - A_t'(w) \), we have

\[
-2\langle A_t(w) - A_t'(w), P_t(w)\zeta_{ta} - P_c(w)v_{ca} \rangle \bigg/ R_n(w) + o(1)
\]

Furthermore, note that

\[
\sup_{w \in H_n} \left\{ \|P_t'(w)\zeta_{ta}\|^2 - \sigma_t^2 tr\left(P_t'(w)P_t(w)\right) \right\} - 2\langle A_t(w), P_t(w)\zeta_{ta} \rangle
\]

\[
+ \left\{ \|P_c(w)v_{ca}\|^2 - \sigma_c^2 tr\left(P_c'(w)P_c(w)\right) \right\} - 2\langle A_c(w), P_c(w)\zeta_{ta} \rangle
\]

\[
+ 2\langle A_t(w), P_c(w)v_{ca} \rangle + 2\langle A_c(w), P_t(w)\zeta_{ta} \rangle
\]

\[
-2\langle A_c(w), P_c(w)v_{ca} \rangle \bigg/ R_n(w) + o(1). \tag{A.26}
\]

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 \( \lambda \). Following [Wan et al. (2010)], we have

\[
\bar{\lambda}(P_t^{(k)}) = \bar{\lambda}(U_t^{(k)}(U_{ta}^{(k)}U_{ta}^{(k)})^{-1}U_{ta}^{(k)\prime})
\]

\[
= \lambda^{1/2}(U_t^{(k)}(U_{ta}^{(k)}U_{ta}^{(k)})^{-1}U_{ta}^{(k)\prime}U_t^{(k)}(U_{ta}^{(k)})^{-1}U_t^{(k)\prime})
\]

\[
= \lambda(U_t^{(k)}(U_{ta}^{(k)}U_{ta}^{(k)})^{-1}U_t^{(k)\prime})
\]

\[
= \bar{\lambda}(U_t^{(k)}(U_{ta}^{(k)}U_{ta}^{(k)}) + F)^{-1}U_t^{(k)\prime} \quad (\exists F \geq 0)
\]

\[
\leq 1.
\]

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

\[
tr\{P_t'(w)P_t(w^*)P_t'(w)\} \leq \bar{\lambda}^2(P_t(w^*))tr\{P_t(w)P_t'(w)\}
\]

\[
\leq tr\{P_t(w)P_t'(w)\}.
\]

Furthermore, note that

\[
\bar{\lambda}(P_c(w^*)) \leq \bar{\lambda}(P_c(w^*)) + \bar{\lambda}(P_c(w^*) - P_c(w^*) - P_c(w^*))
\]

\[
\leq 1 + \max_{1 \leq k \leq K} \lambda^{1/2} \left\{ (U_t^{(k)} - U_c^{(k)})(U_{ca}^{(k)}U_{ca}^{(k)})^{-1}U_{ca}^{(k)\prime} \right\}
\]

\[
\leq 1 + \max_{1 \leq k \leq K} \lambda^{1/2} \left\{ (U_t^{(k)} - U_c^{(k)})(U_{ca}^{(k)}U_{ca}^{(k)})^{-1}U_{ca}^{(k)\prime} \right\}
\]

\[
\leq 1 + \max_{1 \leq k \leq K} \lambda^{1/2} \left\{ n_c^{-1}U_t^{(k)\prime}U_{ca}^{(k)\prime} \right\}
\]

\[
\leq 1 + C_0 \max_{1 \leq k \leq K} \lambda^{1/2} \left\{ n_c^{-1}(U_t^{(k)} - U_c^{(k)})'U_t^{(k)} - U_c^{(k)} \right\}
\]

\[
\leq 1 + C_0 \max_{1 \leq k \leq K} tr^{1/2} \left\{ n_c^{-1}(U_t^{(k)} - U_c^{(k)})'U_t^{(k)} - U_c^{(k)} \right\}
\]
\[
= 1 + C_0 \max_{1 \leq k \leq K} \left[ n_c^{-1} n_2 l_k O \left( \frac{\log n}{n} \right)^{2/p} \right]^{1/2} \\
= 1 + C_0 \max_{1 \leq k \leq K} \left[ \frac{p}{\log n} O \left( \frac{\log n}{n} \right)^{2/p} \right]^{1/2} \\
= 1 + o(1).
\]

Then
\[
P \left\{ \sup_{w \in H_n} \left\| \left\| P_t(w) \xi_{ta} \right\| - \sigma_t^2 \text{tr} \left( P_t'(w) P_t(w) \right) \right\| / R_n(w) > \delta \right\} \\
\leq P \left\{ \sup_{w \in H_n} \sum_{j=1}^{K} \sum_{k=1}^{K} w_j w_k \left| \xi'_{ta} P_t'(w) P_t(w) \xi_{ta} - \sigma_t^2 \text{tr} \left( P_t'(w) P_t(w) \right) \right| > \delta \xi_n \right\} \\
\leq P \left\{ \max_{1 \leq j \leq K} \max_{1 \leq k \leq K} \left| \xi'_{ta} P_t(j) P_t(k) \xi_{ta} - \sigma_t^2 \text{tr} \left( P_t(j) P_t(k) \right) \right| > \delta \xi_n \right\} \\
\leq \sum_{j=1}^{K} \sum_{k=1}^{K} P \left\{ \left\| \left\langle \xi_{ta}, P_t(w_j^0) P_t(w_k^0) \xi_{ta} \right\rangle - \sigma_t^2 \text{tr} \left( P_t(w_j^0) P_t(w_k^0) \right) \right\| > \delta \xi_n \right\} \\
\leq \sum_{j=1}^{K} \sum_{k=1}^{K} E \left( \left\| \left\langle \xi_{ta}, P_t(w_j^0) P_t(w_k^0) \xi_{ta} \right\rangle - \sigma_t^2 \text{tr} \left( P_t(w_j^0) P_t(w_k^0) \right) \right\| \right)^{2G} / \delta^{2G} \\
\leq C_7 \delta^{-2G} \xi_n^{-2G} \sum_{j=1}^{K} \sum_{k=1}^{K} \left( \text{tr} \left( P_t(w_j^0) P_t(w_k^0) \right) \right)^G \\
\leq C_7 \delta^{-2G} \xi_n^{-2G} \sum_{j=1}^{K} \sum_{k=1}^{K} \left( \text{tr} \left( P_t(w_j^0) P_t(w_k^0) \right) \right)^G \\
\leq C_7 \delta^{-2G} \xi_n^{-2G} \sum_{k=1}^{K} \left( R_n(w_k^0) \right)^G \\
\rightarrow 0
\]

and
\[
P \left\{ \sup_{w \in H_n} \left\| \left\| P_t(w) v_{ca} \right\| - \sigma_c^2 \text{tr} \left( P_t'(w) P_t(w) \right) \right\| / R_n(w) > \delta \right\} \\
\leq P \left\{ \sup_{w \in H_n} \sum_{j=1}^{K} \sum_{k=1}^{K} w_j w_k \left| v_{ca} P_t(j) P_t(k) v_{ca} - \sigma_c^2 \text{tr} \left( P_t(j) P_t(k) \right) \right| > \delta \xi_n \right\} \\
\leq P \left\{ \max_{1 \leq j \leq K} \max_{1 \leq k \leq K} \left| v_{ca} P_t(j) P_t(k) v_{ca} - \sigma_c^2 \text{tr} \left( P_t(j) P_t(k) \right) \right| > \delta \xi_n \right\} \\
\leq C_8 \delta^{-2G} \xi_n^{-2G} \sum_{j=1}^{K} \sum_{k=1}^{K} \left( \text{tr} \left( P_t(w_j^0) P_t(w_k^0) \right) \right)^G \\
\rightarrow 0
\]
Additionally, we can show that
\[
\leq C_g \delta^{-2G} \xi_n^{-2G} \left( \lambda(P_t^j(w^0_t) P_t(w^0)) \sum_{j=1}^K \sum_{k=1}^K \left[ \text{tr} \left( P_t^j(w^0_k) P_t(w^0_k) \right) \right]^2 \right)^G
\]
\[
\leq C'_g \delta^{-2G} \xi_n^{-2G} K \sum_{k=1}^K (R_{n_k}(w^0_k))^G
\]
\[
\to 0.
\]

Similarly, we have
\[
P \left\{ \sup_{w \in H_n} \left| \langle A_t(w), P_t(w) \zeta_{ta} \rangle \right| / R_{n}(w) > \delta \right\}
\]
\[
\leq C_{10} \delta^{-2G} \xi_n^{-2G} \sum_{j=1}^K \sum_{k=1}^K \lambda(P_t^j)^{P_t^j} \left\| A_t(w^0_k) \right\| 2^G
\]
\[
\leq C'_{10} \delta^{-2G} \xi_n^{-2G} K \sum_{k=1}^K (R_{n_k}(w^0_k))^G \to 0,
\]
\[
P\left\{ \sup_{w \in H_n} \left| \langle A_\varepsilon(w), P_t(w)\zeta_{ta} \rangle / R_n(w) > \delta \right| \right\} - C_{11} \xi_n^{-2G} \sum_{j=1}^{K} \sum_{k=1}^{K} \lambda(P_t^{(j)} P_t^{(j)}) \| A_\varepsilon(w_0^k) \|^{2G} \\
\leq C_{11} \xi_n^{-2G} \sum_{k=1}^{K} (R_\varepsilon(w_k^0))^G \to 0,
\]

\[
P\left\{ \sup_{w \in H_n} \left| \langle A_\varepsilon(w), P_c(w)\nu_{ca} \rangle / R_n(w) > \delta \right| \right\} - C_{12} \xi_n^{-2G} \sum_{j=1}^{K} \sum_{k=1}^{K} \lambda(P_c^{(j)} P_c^{(j)}) \| A_\varepsilon(w_0^k) \|^{2G} \\
\leq C_{12} \xi_n^{-2G} \sum_{k=1}^{K} (R_n(w_k^0))^G \to 0
\]

and
\[
P\left\{ \sup_{w \in H_n} \left| \langle P_t(w)\zeta_{ta}, P_c(w)\nu_{ca} \rangle / R_n(w) > \delta \right| \right\} - P\left\{ \sup_{w \in H_n} \sum_{j=1}^{K} \sum_{k=1}^{K} w_j w_k \left| P_t^{(j)} P_c^{(k)} \nu_{ca} \right| > \delta \xi_n \right\} \\
\leq P\left\{ \max_{1 \leq j \leq K} \max_{1 \leq k \leq K} \left| P_t^{(j)} P_c^{(k)} \nu_{ca} \right| > \delta \xi_n \right\} \\
\leq \sum_{j=1}^{K} \sum_{k=1}^{K} P\left\{ \left| P_t^{(j)} P_c^{(k)} \nu_{ca} \right| > \delta \xi_n \right\} \\
\leq \sum_{j=1}^{K} \sum_{k=1}^{K} \frac{E \left| P_t^{(j)} P_c^{(k)} \nu_{ca} \right|^2}{\delta^2 \xi_n^2} \\
\leq \sum_{j=1}^{K} \sum_{k=1}^{K} \frac{E \left| P_t^{(j)} \zeta_{ta} \right|^2 E \left| P_c^{(k)} \nu_{ca} \right|^2}{\delta^2 \xi_n^2} \\
\leq C_{13} \xi_n^{-2} \sum_{j=1}^{K} \sum_{k=1}^{K} \text{tr} \left| P_t(w_j^0) P_c(w_k^0) \right|^2 \\
\leq C_{13} \xi_n^{-2} \sum_{j=1}^{K} \sum_{k=1}^{K} \text{tr} \left( P_t(w_j^0) P_c(w_j^0) \right) \text{tr} \left( P_c(w_k^0) P_c(w_k^0) \right) \\
\leq C_{13} \xi_n^{-2} K \sum_{k=1}^{K} (R_n(w_k^0)) \to 0.
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

$$
sup_{w \in H_n} \left| \frac{\tilde{B}(w)}{R_n(w)} \right|
$$

we can see that

$$
sup_{w \in H_n} \frac{|L_n(w) - \tilde{B}(w)|}{R_n(w)} \leq \sup_{w \in H_n} \left( 2\frac{||B(w)||^2}{R_n(w)} + \sup_{w \in H_n} \frac{||B(w)||^2}{R_n(w)} \to_p 0, \right. \quad (A.27)
$$

we can see that

$$
sup_{w \in H_n} \frac{|L_n(w) - R_n(w) - \tilde{B}(w) + \tilde{B}(w)|}{R_n(w)} \to_p 0,
$$

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

$$
\tilde{C}_n(w) = \left| \tilde{Y} - \tilde{\Delta}(w) \right|^2 + 2tr(P_{\tilde{c}_0}(w))\tilde{\sigma}_{\tilde{c}^2_{K^*}} + 2tr(P_{\tilde{c}_0}(w))\tilde{\sigma}_{\tilde{c}^2_{K^*}}
$$

$$
= C_n(w) + 2tr(P_{\tilde{c}_0}(w))(\tilde{\sigma}_{\tilde{c}^2_{K^*}} - \sigma_{\tilde{c}^2_{K^*}}) + 2tr(P_{\tilde{c}_0}(w))(\tilde{\sigma}_{\tilde{c}^2_{K^*}} - \sigma_{\tilde{c}^2_{K^*}}). \quad (A.28)
$$

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

$$
sup_{w \in H_n} tr(P_{\tilde{c}_0}(w)) \left| \tilde{\sigma}_{\tilde{c}^2_{K^*}} - \sigma_{\tilde{c}^2_{K^*}} \right| / R_n(w) \to_p 0 \quad (A.29)
$$

and

$$
sup_{w \in H_n} tr(P_{\tilde{c}_0}(w)) \left| \tilde{\sigma}_{\tilde{c}^2_{K^*}} - \sigma_{\tilde{c}^2_{K^*}} \right| / R_n(w) \to_p 0. \quad (A.30)
$$

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

It should be mentioned that our assumption of $U$ being non-stochastic instead of allowing it to be stochastic has no impact on our proof, because all of our technical conditions related to $U$ hold almost surely. Let $r_{\tilde{c}} = tr(P_{\tilde{c}_0}(w))$.

Then we have

$$
r_{\tilde{c}} \leq \tilde{\lambda}(P_{\tilde{c}_0}(w)) l_{K^*} = (1 + o_p(1))l_{K^*}. \quad (A.31)
$$
and
\[
\sup_{w \in H_n} \left| \frac{\text{tr}\left[ \mathbf{P}_{z0}(w) \right] - \mathbf{I}}{R_n(w)} \right| \leq \frac{r_n}{\xi_n} \left| \frac{\hat{\sigma}_c^2}{\sigma_c^2} \right| + \frac{r_n}{\xi_n} \left| \frac{Y_{\epsilon \eta}(I - \mathbf{P}_{ca}(K^*))Y_{\epsilon \eta}}{n_c - l_{K^*}} - \sigma_c^2 \right|
\]
\[
\leq \frac{r_n}{n_c - l_{K^*}} \frac{f'_{ca}(I - \mathbf{P}_{ca}(K^*))f_{ca}}{\xi_n} + \frac{2r_n}{\xi_n} \left| \frac{f'_{ca}(I - \mathbf{P}_{ca}(K^*))v_{ca}}{\sigma_c^2} \right| \rightarrow 0
\]
\[
(A.32)
\]

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
\[
f'_{ca}(I - \mathbf{P}_{ca}(K^*))f_{ca} \rightarrow 0.
\]
\[
(A.33)
\]

Then from (24), (26), (A.31) and (A.33), it is seen that
\[
\frac{r_n}{n_c - l_{K^*}} \frac{f'_{ca}(I - \mathbf{P}_{ca}(K^*))f_{ca}}{\xi_n} \left| \frac{\bar{f}'_{ca}(I - \mathbf{P}_{ca}(K^*))v_{ca}}{\sigma_c^2} \right| \left| \frac{\bar{f}_ca f_{ca}}{\sigma_c^2} \right| \rightarrow 0
\]
\[
(A.34)
\]

and
\[
P\left\{ \frac{2r_n}{\xi_n} \left| \frac{f'_{ca}(I - \mathbf{P}_{ca}(K^*))v_{ca}}{\sigma_c^2} \right| > \delta \right\}
\leq E \left[ \left| \frac{f'_{ca}(I - \mathbf{P}_{ca}(K^*))v_{ca}}{\sigma_c^2} \right| \right] \leq C_{14} r_n^2 \frac{f'_{ca}(I - \mathbf{P}_{ca}(K^*))f_{ca}}{\delta^2 \xi_n^2 (n_c - l_{K^*})^2} \rightarrow 0.
\]
\[
(A.35)
\]

By (A.35), it can be seen the second term on the right hand side of (A.32) tends to zero in probability. From
\[
P\left\{ \frac{r_n}{\xi_n} \left| \frac{v'_{ca}(I - \mathbf{P}_{ca}(K^*))v_{ca}}{\sigma_c^2} \right| > \delta \right\}
\leq E \left[ \left| \frac{v'_{ca}(I - \mathbf{P}_{ca}(K^*))v_{ca}}{\sigma_c^2} \right| \right] \leq C_{15} r_n^2 \frac{r_n^2}{\delta^2 \xi_n^2 (n_c - l_{K^*})^2} \rightarrow 0.
\]
\[
(A.36)
\]

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

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

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 \( \zeta_i \) 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 \( np^{-13/6} \xi_n = o(1) \). Hence Condition (14) holds if \( p = O((\log n)^{1/5}) \).

A.5 Results of Design 2

![Graphs showing results of simulations under Design 2 with \( n = 100 \).](image)

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

