

# Supplement to “Information theoretic approach to high-dimensional multiplicative models: Stochastic discount factor and treatment effect”

(Quantitative Economics, Vol. 13, No. 1, January 2022, 63–94)

CHEN QIU

Department of Economics, Cornell University

TAISUKE OTSU

Department of Economics, London School of Economics

## APPENDIX A: PROOFS FOR THE LOW-DIMENSIONAL CASE

Recall  $g_n(X) = \mathbb{E}[g(X)g(X)'\mathbb{I}\{X \in \mathcal{X}_n\}]^{-1/2}g(X)\mathbb{I}\{X \in \mathcal{X}_n\}$  and  $r_n(X) = \mathbb{E}[g(X)g(X)'\mathbb{I}\{X \in \mathcal{X}_n\}]^{-1/2}r(X)\mathbb{I}\{X \in \mathcal{X}_n\}$ . Define  $\tilde{\lambda} = \arg \min_{\lambda} \mathbb{E}_n[\phi_*(\lambda'g_n(X)) - \lambda'r_n(X)]$ .

### A.1 Lemmas

**LEMMA 1.** *Let  $f(x) = (f_1(x), \dots, f_K(x))'$  be a  $K$ -dimensional vector of functions, and  $M_q = \max_{1 \leq j \leq K} \{\mathbb{E}|f_j(X)|^q\}^{1/q}$ . Suppose  $\{X_i\}_{i=1}^n$  is  $\alpha$ -mixing with mixing coefficient  $\{\alpha_m\}_{m \in \mathbb{N}}$  satisfying  $KM_2(M_2 + M_q \sum_{m=1}^n \alpha_m^{1/2-1/q})/n \rightarrow 0$  for some  $q \in (2, \infty)$ . Then*

$$|\mathbb{E}_n[f(X)] - \mathbb{E}[f(X)]| = O_p\left(\sqrt{\frac{KM_2}{n} \left(M_2 + M_q \sum_{m=1}^n \alpha_m^{1/2-1/q}\right)}\right).$$

**LEMMA 2.** *Suppose Conditions D, S, and I hold true. Then*

- (i) *for all  $x \in \mathcal{X}$  and  $n$  large enough,  $\lambda'_b g_n(x) \in \mathcal{C}$ , where  $\mathcal{C}$  is a compact set in  $(\phi^{(1)}(0), \phi^{(1)}(+\infty))$ ,*
- (ii)  $\sup_{x \in \mathcal{X}_n} |\omega_0(x) - \phi_*^{(1)}(\lambda'_b g_n(x))| = O(\eta_{K,n})$ .

**LEMMA 3.** *Suppose the conditions for Theorem 1 hold true. Then*

- (i) *if we additionally assume that  $\{X_i\}_{i=1}^n$  is iid and  $\zeta_{K,n}^2 \log K/n \rightarrow 0$ , then  $|\mathbb{E}_n \times [g_n(X)g_n(X)'] - I| = O_p(\sqrt{\zeta_{K,n}^2 \log K/n})$ , and thus  $\lambda_{\min}(\mathbb{E}_n[g_n(X)g_n(X)'])$  is bounded away from zero and from above with probability approaching to one,*
- (ii)  $|\mathbb{E}_n[r_n(X) - \omega_0(X)g_n(X)]| = O_p(\sqrt{K\mu_{K,n}/n})$ ,

---

Chen Qiu: [cq62@cornell.edu](mailto:cq62@cornell.edu)

Taisuke Otsu: [t.otsu@lse.ac.uk](mailto:t.otsu@lse.ac.uk)

$$(iii) \quad |\mathbb{E}_n[\{\omega_0(X) - \phi_*^{(1)}(\lambda'_b g_n(X))\}g_n(X)]| = O_p(B_{K,n}),$$

$$(iv) \quad |\tilde{\lambda} - \lambda_b| = O_p(\sqrt{K\mu_{K,n}/n} + B_{K,n}).$$

*Proof of Lemma 1*

Let  $W(X) = f(X) - \mathbb{E}[f(X)]$ . Note that

$$\mathbb{E}[|\mathbb{E}_n[W(X_i)]|^2] = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^K \mathbb{E}[W_j(X_i)^2] + \frac{1}{n^2} \sum_{i \neq l}^n \sum_{j=1}^K \mathbb{E}[W_j(X_i)W_j(X_l)].$$

The first term is bounded as  $\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^K \mathbb{E}[W_j^2(X_i)] \leq KM_2^2/n$ . For the second term, Hall and Heyde (2014, Corollary A.2) implies

$$|\mathbb{E}[W_j(X_i)W_j(X_l)]| \lesssim \{\mathbb{E}[|W_j(X_i)|^q]\}^{1/q} \sqrt{\mathbb{E}[W_j(X_l)^2]} \alpha_{i-l}^{1/2-1/q} \leq M_q M_2 \alpha_{i-l}^{1/2-1/q},$$

and thus  $\frac{1}{n^2} \sum_{i \neq l}^n \sum_{j=1}^K \mathbb{E}[W_j(X_i)W_j(X_l)] \lesssim KM_q M_2 \sum_{m=1}^n \alpha_m^{1/2-1/q}$ . Therefore, the conclusion follows by Markov's inequality.

*Proof of Lemma 2(i)*

By boundedness and positivity of  $\omega_0$  (Condition D(3)) and continuous differentiability and strict convexity of  $[\phi_*^{(1)}]^{-1}(\cdot)$  on  $(0, +\infty)$  (Condition D(4), since  $[\phi_*^{(1)}]^{-1}(\cdot) = \phi^{(1)}(\cdot)$  on  $(0, +\infty)$ ), both  $\phi^{(1)}(0) < \underline{\gamma} = \inf_{x \in \mathcal{X}} [\phi_*^{(1)}]^{-1}(\omega_0(x))$  and  $\bar{\gamma} = \sup_{x \in \mathcal{X}} [\phi_*^{(1)}]^{-1}(\omega_0(x))$  are finite. Thus, by (15) in Condition S, there exists  $C_1 > 0$  such that

$$\lambda'_b g_n(x) \in [\underline{\gamma} - C_1 \eta_{K,n}, \bar{\gamma} + C_1 \eta_{K,n}], \quad (37)$$

for all  $x \in \mathcal{X}_n$ . The conclusion holds for all  $x \in \mathcal{X}$  by the requirement  $\eta_{K,n} \rightarrow 0$  and  $\phi^{(1)}(0) < 0$  from Condition D(4).

*Proof of Lemma 2(ii)*

Note that (37) also guarantees

$$\begin{aligned} \omega_0(x) - \phi_*^{(1)}(\lambda'_b g_n(x)) &\in [\phi_*^{(1)}(\lambda'_b g_n(x) - C_1 \eta_{K,n}) - \phi_*^{(1)}(\lambda'_b g_n(x)), \\ &\quad \phi_*^{(1)}(\lambda'_b g_n(x) + C_1 \eta_{K,n}) - \phi_*^{(1)}(\lambda'_b g_n(x))], \end{aligned}$$

for all  $x \in \mathcal{X}_n$  and  $n$  large enough. By applying the mean value theorem to the upper and lower bounds under Condition I, there exist  $c_1, c_2 > 0$  such that

$$\begin{aligned} \phi_*^{(1)}(\lambda'_b g_n(x) + C_1 \eta_{K,n}) - \phi_*^{(1)}(\lambda'_b g_n(x)) &\leq c_1 C_1 \eta_{K,n}, \\ \phi_*^{(1)}(\lambda'_b g_n(x) - C_1 \eta_{K,n}) - \phi_*^{(1)}(\lambda'_b g_n(x)) &\geq -c_2 C_1 \eta_{K,n}, \end{aligned}$$

for all  $x \in \mathcal{X}_n$  and  $n$  large enough. Combining these results, the conclusion follows.

*Proof of Lemma 3(i)*

This follows directly from Belloni et al. (2015, Lemma 6.2) or Chen and Christensen (2015, Lemma 2.1).

*Proof of Lemma 3(ii)*

Let  $f(x) = r_n(x) - \omega_0(x)g_n(x)$ . By (1) and the Cauchy–Schwarz inequality, we have

$$\begin{aligned} |\mathbb{E}[f(X)]| &\lesssim |\mathbb{E}[\{\omega_0(X)g(X) - r(X)\}\mathbb{I}\{X \notin \mathcal{X}_n\}]|, \\ &\leq \sqrt{\mathbb{E}[\{\omega_0(X)g(X) - r(X)\}^2]}\sqrt{\mathbb{P}\{X \notin \mathcal{X}_n\}} = o(\sqrt{K/n}), \end{aligned} \quad (38)$$

where the equality follows from Condition S. Condition S guarantees  $\max_{1 \leq j \leq K} \{\mathbb{E} \times \{|f_j(X)|^q\}^{1/q} \lesssim M_{K,n}$ . Thus, Lemma 1 implies

$$|\mathbb{E}_n[f(X)] - \mathbb{E}[f(X)]| = O_p(\sqrt{K\mu_{K,n}/n}). \quad (39)$$

The conclusion follows by (38) and (39).

*Proof of Lemma 3(iii)*

Let

$$\xi(X) = \{\omega_0(X) - \phi_*^{(1)}(\lambda'_b g_n(X))\}, \quad \hat{\rho} = (\mathbb{E}_n[g_n(X)g_n(X)'])^{-1}\mathbb{E}_n[g_n(X)\xi(X)].$$

By the assumption  $|\mathbb{E}_n[g_n(X)g_n(X)'] - I| = o_p(1)$ , it holds  $(\mathbb{E}_n[g_n(X)g_n(X)'])^{-1} = O_p(1)$ , and then

$$|\mathbb{E}_n[g_n(X)\xi(X)]| \leq |\mathbb{E}_n[g_n(X)g_n(X)']||\hat{\rho}| \lesssim |\hat{\rho}| \lesssim \sqrt{\mathbb{E}_n[(\hat{\rho}'g_n(X))^2]}, \quad (40)$$

with probability approaching one, where the last inequality follows from Condition S. Since  $\hat{\rho}$  is the empirical projection coefficient from  $\xi(X)$  on  $g_n(X)$ , we have

$$\mathbb{E}_n[(\hat{\rho}'g_n(X))^2] \leq \{\mathbb{E}_n[\xi(X)^2] - \mathbb{E}[\xi(X)^2]\} + \mathbb{E}[\xi(X)^2] = O_p(B_{K,n}^2), \quad (41)$$

where the equality follows from (16) in Condition S and Lemma 1 (note that  $\mathbb{E}[|\xi(X)|^q] \lesssim s_{K,n}^{2/q}$  under Conditions D and S). The conclusion follows from (40) and (41).

*Proof of Lemma 3(iv)*

Recall that  $\hat{\omega}(X) = \phi_*^{(1)}(\hat{\lambda}'g(X)\mathbb{I}\{X \in \mathcal{X}_n\}) = \phi_*^{(1)}(\tilde{\lambda}'g_n(X))$ , where  $\tilde{\lambda} = \arg \max_{\lambda} \hat{Q}(\lambda)$  and

$$\hat{Q}(\lambda) = \lambda' \mathbb{E}_n[r_n(X)] - \mathbb{E}_n[\phi_*(\lambda'g_n(X))].$$

By Condition D,  $\hat{Q}(\lambda)$  is concave. Let  $\hat{Q}^{(1)}(\lambda)$  and  $\hat{Q}^{(2)}(\lambda)$  be the first and second derivatives of  $\hat{Q}(\lambda)$ , respectively, if they exist. The proof is split into several steps.

**Step 1:** Show  $\hat{Q}^{(1)}(\lambda_b) = O_p(\delta_n)$ , where  $\delta_n = \sqrt{K\mu_{K,n}/n} + B_{K,n}$ . Since  $\hat{Q}^{(1)}(\lambda_b) = \mathbb{E}_n[r_n(X) - \phi_*^{(1)}(\lambda'_b g_n(X))g_n(X)]$ , the triangle inequality yields

$$|\hat{Q}^{(1)}(\lambda_b)| \leq |\mathbb{E}_n[r_n(X) - \omega_0(X)g_n(X)]| + |\mathbb{E}_n[\{\omega_0(X) - \phi_*^{(1)}(\lambda'_b g_n(X))\}g_n(X)]|$$

Thus, Lemma 3(ii) and (iii) imply  $\hat{Q}^{(1)}(\lambda_b) = O_p(\delta_n)$ .

**Step 2:** Show that for any  $C > 0$ , there exists some  $c > 0$  such that

$$\eta_C = \inf_{|\lambda - \lambda_b| \leq C\delta_n, x \in \mathcal{X}} \phi_*^{(2)}(\lambda' g_n(x)) > c.$$

Pick any  $C > 0$ . Since  $\delta_n \zeta_{K,n} = o(1)$ , we have

$$|\lambda' g_n(x)| \leq |\lambda'_b g_n(x)| + |\lambda - \lambda_b| |g_n(x)| \leq |\lambda'_b g_n(x)| + C\delta_n \zeta_{K,n},$$

for all  $\lambda$  satisfying  $|\lambda - \lambda_b| \leq C\delta_n$ . Thus, by Lemma 2(i),  $\lambda' g_n(x)$  lies in some compact set  $\tilde{\mathcal{C}}$  in  $(\phi^{(1)}(0), \phi^{(1)}(+\infty))$  for all  $\lambda$  satisfying  $|\lambda - \lambda_b| \leq C\delta_n$  and  $x \in \mathcal{X}$ . Condition I and the Weierstrass theorem guarantee  $\eta_C > c = \min_{a \in \tilde{\mathcal{C}}} \phi_*^{(2)}(a) > 0$ .

**Step 3:** Show that there exists some  $C^* > 0$  such that  $\hat{Q}(\lambda) < \hat{Q}(\lambda_b)$  with probability approaching one for all  $\lambda$  satisfying  $|\lambda - \lambda_b| = C^*\delta_n$ . Pick any  $\epsilon > 0$ . By Step 1, we can take  $C^* > 0$  such that

$$\mathbb{P}\{|\hat{Q}^{(1)}(\lambda_b)| < cC^*\delta_n/4\} \geq 1 - \epsilon, \quad (42)$$

for all  $n$  large enough, where  $c > 0$  is chosen in Step 2. An expansion of  $\hat{Q}(\lambda)$  around  $\lambda = \lambda_b$  yields

$$\hat{Q}(\lambda) - \hat{Q}(\lambda_b) = \hat{Q}^{(1)}(\lambda_b)'(\lambda - \lambda_b) + \frac{1}{2}(\lambda - \lambda_b)' \hat{Q}^{(2)}(\dot{\lambda})(\lambda - \lambda_b),$$

for some  $\dot{\lambda}$  on the line joining  $\lambda$  and  $\lambda_b$ . By Step 2,

$$\hat{Q}^{(2)}(\dot{\lambda}) = -\mathbb{E}_n[\phi_*^{(2)}(\dot{\lambda}' g_n(X))g_n(X)g_n(X)'] \leq_{\text{psd}} -c\mathbb{E}_n[g_n(X)g_n(X)'],$$

and Condition S(1) implies

$$\frac{1}{2}(\lambda - \lambda_b)' \hat{Q}^{(2)}(\dot{\lambda})(\lambda - \lambda_b) \leq -\frac{c}{4}|\lambda - \lambda_b|^2,$$

with probability approaching one. Combining these results, for all  $\lambda$  satisfying  $|\lambda - \lambda_b| = C^*\delta_n$ ,

$$\hat{Q}(\lambda) - \hat{Q}(\lambda_b) \leq |\hat{Q}^{(1)}(\lambda_b)| |\lambda - \lambda_b| - \frac{c}{4}|\lambda - \lambda_b|^2 \leq \left( |\hat{Q}^{(1)}(\lambda_b)| - \frac{cC^*\delta_n}{4} \right) |\lambda - \lambda_b|.$$

Thus, (42) implies that  $\hat{Q}(\lambda) < \hat{Q}(\lambda_b)$  with probability approaching one.

**Step 4:** By continuity of  $\hat{Q}(\lambda)$ , it has a maximum on the compact set  $\{\lambda : |\lambda - \lambda_b| \leq C^*\delta_n\}$ . By Step 3, the maximum  $\tilde{\lambda}_{C^*}$  on set  $\{\lambda : |\lambda - \lambda_b| \leq C^*\delta_n\}$  must satisfy  $|\tilde{\lambda}_{C^*} - \lambda_b| < C^*\delta_n$ . By concavity of  $\hat{Q}(\lambda)$ ,  $\tilde{\lambda}_{C^*}$  also maximizes  $\hat{Q}(\lambda)$  over  $\mathbb{R}^k$ . The conclusion follows by the same argument used at the end of the proof of Newey and McFadden (1994, Theorem 2.7).

## A.2 Proof of Theorem 1

## Proof of (17)

Let  $\omega_b(x) = \phi_*^{(1)}(\lambda'_b g_n(x))$ . Pick any  $C > 0$ . From Step 2 in the proof of Lemma 3(iv),  $\lambda'_b g_n(x)$  lies in some compact set  $\tilde{\mathcal{C}}$  in  $(\phi^{(1)}(0), \phi^{(1)}(+\infty))$  for all  $x \in \mathcal{X}$  and  $\lambda$  satisfying  $|\lambda - \lambda_b| \leq C\delta_n$ . Let  $\mathcal{E}_n$  be the event that  $\tilde{\lambda}'_x g_n(x) \in \tilde{\mathcal{C}}$  for all  $x \in \mathcal{X}$ . Lemma 3(iv) guarantees  $\mathbb{P}\{\mathcal{E}_n\} \rightarrow 1$ . On event  $\mathcal{E}_n$ , an expansion around  $\tilde{\lambda} = \lambda_b$  yields

$$\hat{\omega}(x) - \omega_b(x) = \phi_*^{(2)}(\tilde{\lambda}'_x g_n(x))(\tilde{\lambda} - \lambda_b)' g_n(x), \quad (43)$$

where  $\tilde{\lambda}_x$  is a point on the line joining  $\tilde{\lambda}$  and  $\lambda_b$ , and  $\tilde{\lambda}'_x g_n(x) \in \tilde{\mathcal{C}}$  for all  $x \in \mathcal{X}$ . The Weierstrass theorem and Condition I imply

$$\sup_{|\lambda - \lambda_b| \leq C\delta_n, x \in \mathcal{X}} \phi_*^{(2)}(\lambda'_x g_n(x)) < C_1 < \infty, \quad (44)$$

for some  $C_1 > 0$ . Furthermore, observe that

$$\begin{aligned} \mathbb{E}_n[\{\hat{\omega}(X) - \omega_b(X)\}^2] &= (\tilde{\lambda} - \lambda_b)' \mathbb{E}_n[\{\phi_*^{(2)}(\tilde{\lambda}'_X g_n(X))\}^2 g_n(X) g_n(X)'] (\tilde{\lambda} - \lambda_b) \\ &\leq C_1 |\tilde{\lambda} - \lambda_b|^2 |\mathbb{E}_n[g_n(X) g_n(X)']| \\ &= O_p(|\tilde{\lambda} - \lambda_b|^2), \end{aligned} \quad (45)$$

where the inequality follows from (44) and  $\mathbb{P}\{\mathcal{E}_n\} \rightarrow 1$ , and the second equality follows from Condition S and Lemma 3(iv). Now, the same argument in the proof of Lemma 3(iii) for (41) yields

$$\mathbb{E}_n[\{\omega_b(X) - \omega_0(X)\}^2] = O_p(B_{K,n}^2). \quad (46)$$

The conclusion follows by (45), (46), and the triangle inequality.

Proof of  $\hat{\theta} \xrightarrow{P} \theta_0$ 

Observe that

$$\begin{aligned} |\hat{\theta} - \theta_0| &\leq |\mathbb{E}_n[\hat{\omega}(X)h(X, Y)] - \mathbb{E}_n[\omega_0(X)h(X, Y)]| + |\mathbb{E}_n[\omega_0(X)h(X, Y)] \\ &\quad - \mathbb{E}[\omega_0(X)h(X, Y)]| \\ &\leq \sqrt{\mathbb{E}_n[\{\hat{\omega}(X) - \omega_0(X)\}^2]} \sqrt{\mathbb{E}_n[h(X, Y)^2]} + |\mathbb{E}_n[\omega_0(X)h(X, Y)] \\ &\quad - \mathbb{E}[\omega_0(X)h(X, Y)]| \\ &= O_p(\sqrt{K\mu_{K,n}/n} + B_{K,n}) + o_p(1), \end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from the Cauchy–Schwarz inequality, and the final equality follows from the law of large numbers (under Condition D) for stationary and ergodic processes and (17) in Theorem 1.

*Proof of (18)*

By the triangle inequality,

$$\sup_{x \in \mathcal{X}_n} |\hat{\omega}(x) - \omega_0(x)| \leq \sup_{x \in \mathcal{X}_n} |\hat{\omega}(x) - \omega_b(x)| + \sup_{x \in \mathcal{X}_n} |\omega_b(x) - \omega_0(x)|.$$

From the proof of (17), it is easy to see that  $\sup_{x \in \mathcal{X}_n} |\hat{\omega}(x) - \omega_b(x)| = O_p(\zeta_{K,n}(\sqrt{K\mu_{K,n}/n} + B_{K,n}))$ . Thus, the conclusion follows by Lemma 2(ii).

A.3 *Proof of Theorem 2*

Let

$$\begin{aligned} h_i &= h(X_i, Y_i), & h_i^X &= \mathbb{E}[h_i | X_i], & \omega_{0i} &= \omega_0(X_i), & g_{ni} &= g_n(X_i), \\ \omega_{bi} &= \phi_*^{(1)}(\lambda'_b g_{ni}), & \hat{\omega}_i &= \phi_*^{(1)}(\tilde{\lambda}' g_{ni}), & r_{ni} &= r_n(X_i), & r_i^h &= r^h(X_i). \end{aligned} \quad (47)$$

By an expansion of  $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \phi_*^{(1)}(\tilde{\lambda}' g_{ni}) h_i$  around  $\tilde{\lambda} = \lambda_b$ , we decompose

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{0i} h_i - \theta_0) + T_1 + T_2 + T_3 + T_4,$$

where

$$\begin{aligned} T_1 &= \mathbb{E}[\phi_*^{(2)}(\lambda'_b g_{ni}) h_i g'_{ni}] \sqrt{n}(\tilde{\lambda} - \lambda_b), \\ T_2 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{ \phi_*^{(2)}(\lambda'_b g_{ni}) h_i g'_{ni} - \mathbb{E}[\phi_*^{(2)}(\lambda'_b g_{ni}) h_i g'_{ni}] \} (\tilde{\lambda} - \lambda_b), \\ T_3 &= \frac{1}{2} (\tilde{\lambda} - \lambda_b)' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_*^{(3)}(\lambda'_b g_{ni}) h_i g_{ni} g'_{ni} \right) (\tilde{\lambda} - \lambda_b), \\ T_4 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{bi} h_i - \omega_{0i} h_i), \end{aligned}$$

and  $\tilde{\lambda}$  lies on the line joining  $\tilde{\lambda}$  and  $\lambda_b$ .

First, we consider  $T_2$ . Since Lemma 2(i) and Assumption N imply  $\max_{1 \leq j \leq K} \{\mathbb{E}[|\phi_*^{(2)} \times (\lambda'_b g_n) h g_{nj}|^2]\} \lesssim 1$  and  $\max_{1 \leq j \leq K} \{\mathbb{E}[|\phi_*^{(2)}(\lambda'_b g_w) h g_{nj}|^{q_1}]\}^{1/q_1} \lesssim M_{K,n}$ , Lemma 1 yields

$$\left| \frac{1}{n} \sum_{i=1}^n \{ \phi_*^{(2)}(\lambda'_b g_{ni}) h_i g'_{ni} - \mathbb{E}[\phi_*^{(2)}(\lambda'_b g_{ni}) h_i g'_{ni}] \} \right| = O_p\left(\sqrt{\frac{K\mu_{K,n}}{n}}\right).$$

Thus, the Cauchy–Schwarz inequality and Lemma 3(iv) imply  $T_2 = O_p(\sqrt{K\mu_{K,n}} \times (\sqrt{K\mu_{K,n}/n} + B_{K,n}))$ .

Next, we consider  $T_3$ . The definitions of  $\zeta_{K,n}$  and matrix  $L_2$ -norm, Lemmas 2(i) and 3(iv), and Condition I imply  $|\frac{1}{n} \sum_{i=1}^n \phi_*^{(3)}(\lambda' g_{ni}) h_i g_{ni} g'_{ni}| = O_p(\zeta_{K,n}^2)$ . Thus, the Cauchy-Schwarz inequality and Lemma 3(iv) imply

$$T_3 = O_p(\sqrt{n} \zeta_{K,n}^2 (K \mu_{K,n}/n + B_{K,n}^2)).$$

Third, we consider  $T_4$ . From the proof of Lemma 3(iii) and the law of large numbers, we have  $T_4 = O_p(\sqrt{n} B_{K,n})$ .

We now consider  $T_1$ . By expanding the first-order condition of  $\tilde{\lambda}$ ,

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n \{ \phi_*^{(1)}(\tilde{\lambda}' g_{ni}) g_{ni} - r_{ni} \} \\ &= \frac{1}{n} \sum_{i=1}^n (\omega_{bi} g_{ni} - r_{ni}) + \frac{1}{n} \sum_{i=1}^n \phi_*^{(2)}(\tilde{\lambda}' g_{ni}) g_{ni} g'_{ni} (\tilde{\lambda} - \lambda_b), \end{aligned} \quad (48)$$

where  $\bar{\lambda}$  lies on the line joining  $\tilde{\lambda}$  and  $\lambda_b$ . Let  $\psi = \mathbb{E}[\phi_*^{(2)}(\lambda'_b g_{ni}) h_i g'_{ni}]$ ,  $\Sigma = \mathbb{E}[\phi_*^{(2)}(\lambda'_b g_{ni}) \times g_{ni} g'_{ni}]$ , and  $\bar{\Sigma} = \frac{1}{n} \sum_{i=1}^n \phi_*^{(2)}(\tilde{\lambda}' g_{ni}) g_{ni} g'_{ni}$ . By solving this for  $\tilde{\lambda} - \lambda_b$  and inserting to  $T_1$ , we have

$$T_1 = -\psi \bar{\Sigma}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{bi} g_{ni} - r_{ni}) = T_{11} + T_{12} + T_{13},$$

where

$$\begin{aligned} T_{11} &= -\psi (\bar{\Sigma}^{-1} - \Sigma^{-1}) \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{bi} g_{ni} - r_{ni}), \\ T_{12} &= -\psi \Sigma^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{bi} - \omega_{0i}) g_{ni}, \\ T_{13} &= -\psi \Sigma^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{0i} g_{ni} - r_{ni}). \end{aligned}$$

For  $T_{12}$ , note that

$$|T_{12}| \leq |\psi| \frac{1}{\lambda_{\min}(\Sigma)} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{bi} - \omega_{0i}) g_{ni} \right|.$$

It is easy to see that  $|\psi| = O(\zeta_{K,n})$  due to the definition of  $\zeta_{K,n}$ . Lemma 3(iii) yields  $|\frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{bi} - \omega_{0i}) g_{ni}| = O_p(\sqrt{n} B_{K,n})$ . Since  $\lambda_{\min}(\Sigma)$  is bounded away from zero by Condition D and Lemma 2(i), we have  $T_{12} = O_p(\sqrt{n} \zeta_{K,n} B_{K,n})$ . For  $T_{11}$ , note that (48) implies

$$T_{11} = \sqrt{n} \psi (\bar{\Sigma}^{-1} - \Sigma^{-1}) \bar{\Sigma} (\tilde{\lambda} - \lambda_b) = \sqrt{n} \psi \Sigma^{-1} (\Sigma - \bar{\Sigma}) (\tilde{\lambda} - \lambda_b),$$

which can be bounded as  $|T_{11}| \leq \sqrt{n}|\psi| \frac{1}{\lambda_{\min}(\Sigma)} |\Sigma - \bar{\Sigma}| \cdot |\bar{\lambda} - \lambda_b|$ . By the triangle inequality and Condition N(2),

$$|\Sigma - \bar{\Sigma}| \leq |\mathbb{E}_n[(\phi_*^{(2)}(\bar{\lambda}'g_n) - \phi_*^{(2)}(\lambda'_b g_n))g_n g_n']| + O_p(\Gamma_{K,n}).$$

By an expansion of  $\phi_*^{(2)}(\bar{\lambda}'g_{ni})$  and Lemmas 2(i) and 3(iv), we have  $|\mathbb{E}_n[(\phi_*^{(2)}(\bar{\lambda}'g_n) - \phi_*^{(2)}(\lambda'_b g_n))g_n g_n']| = O_p(\zeta_K^3(\sqrt{K\mu_{K,n}/n} + B_{K,n}))$ . Therefore, we obtain

$$|\Sigma - \bar{\Sigma}| = O_p(\zeta_{K,n}^3(\sqrt{K\mu_{K,n}/n} + B_{K,n}) + \Gamma_{K,n}).$$

Also by  $|\psi| = O(\zeta_{K,n})$  and Lemma 3(iv), we have

$$|T_{11}| = O_p(\sqrt{n}\zeta_{K,n}^4(K\mu_{K,n}/n + B_{K,n}^2) + \sqrt{n}\zeta_{K,n}\Gamma_{K,n}(\sqrt{K\mu_{K,n}/n} + B_{K,n})).$$

Now consider  $T_{13}$ . Note that

$$\begin{aligned} T_{13} &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{0i} h_i^X - r_i^h) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\beta'(\omega_{0i} g_{ni} - r_{ni}) - (\omega_{0i} h_i^X - r_i^h)\} \\ &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{0i} h_i^X - r_i^h) + o_p(1), \end{aligned}$$

where the second equality follows from Lemma 1 and the condition (19).

Combining these results, we obtain

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\omega_{0i} h_i - \theta_0 - (\omega_{0i} h_i^X - r_i^h)\} + O_p(r_n),$$

where  $r_n = (\sqrt{n}(\zeta_{K,n}^4 K\mu_{K,n}/n + \zeta_{K,n} B_{K,n} + \sqrt{K\mu_{K,n}/n} \zeta_{K,n} \Gamma_{K,n}))$ . Since  $r_n \rightarrow 0$  by the assumption, the central limit theorem for  $\alpha$ -mixing processes (e.g., Theorem 0 in Bradley (1985)) yields the conclusion.

#### A.4 Proof of Proposition 2

*Proof of (i)* In this case,  $r(X)$  is a constant vector  $r = \mathbb{E}[\omega_{0i} g_i]$ . We set  $r^h(X)$  as a constant vector  $r^h = \mathbb{E}[\omega_{0i} h_i^X]$ . Observe that

$$\mathbb{E}[\beta'(\omega_{0i} g_{ni} - r_{ni}) - (\omega_{0i} h_i^X - \mathbb{E}[\omega_{0i} h_i^X])]^2 \leq N_1 + N_2 + N_3,$$

where

$$N_1 = \mathbb{E}[\beta'(\omega_{0i} g_{ni} - \mathbb{E}[\omega_{0i} g_{ni}]) - (\omega_{0i} h_i^X - \mathbb{E}[\omega_{0i} h_i^X])]^2,$$

$$N_2 = \mathbb{E}[\beta'(\mathbb{E}[\omega_{0i} g_{ni}] - \mathbb{E}[r_{ni}])]^2,$$

$$N_3 = \mathbb{E}[\beta'(\mathbb{E}[r_{ni}] - r_{ni})]^2.$$



For  $N_1$ ,

$$N_1 \leq \mathbb{E}[\omega_{0i}^2(h_i^X - \beta' g_{ni})^2] \leq \left( \sup_{x \in \mathcal{X}} \frac{\omega_0^2(x)}{\phi_*^{(2)}(\lambda'_b g_n(x))} \right) \mathbb{E}[(\tilde{h}_i - \beta'_p \tilde{g}_{ni})^2],$$

where  $\tilde{h}_i = \sqrt{\phi_*^{(2)}(\lambda'_b g_{ni})} h_i^X$ ,  $\tilde{g}_i = \sqrt{\phi_*^{(2)}(\lambda'_b g_{ni})} g_{ni}$ , and  $\beta_p = \mathbb{E}[\tilde{g}_{ni} \tilde{g}'_{ni}]^{-1} \mathbb{E}[\tilde{g}_{ni} \tilde{h}_i]$ . Since  $\beta_p$  is the projection coefficient that solves  $\min_b \mathbb{E}[(\tilde{h}_i - b' \tilde{g}_{ni})^2]$ , the assumption in (21) guarantees  $N_1 = o(n^{-1})$ . For  $N_2$ , (38) implies  $|\beta| = O(1)$  (because  $\beta$  is a projection coefficient). By (21), we have

$$N_2 \lesssim \mathbb{E}[|\omega_0(X)g(X) - r(X)|^2] \mathbb{P}\{X \notin \mathcal{X}_n\} = o(n^{-1}).$$

For  $N_3$ , the definition of  $r_{ni}$ ,  $|\beta| = O(1)$ , and (21) imply

$$N_3 = \mathbb{E}[\beta'(r_{ni} - \mathbb{E}[r_{ni}])]^2 \lesssim |\beta|^2 K \mathbb{P}\{X \in \mathcal{X}_n\} \mathbb{P}\{X \notin \mathcal{X}_n\} = o(n^{-1}).$$

Combining these results, the conclusion follows.

*Proof of (ii)* This follows by a standard projection argument and thus the proof is omitted.

## APPENDIX B: PROOFS FOR HIGH-DIMENSIONAL CASE

### B.1 Proof of Theorem 3

By the mean value theorem, there exists  $t_x \in [0, 1]$  such that

$$\hat{\omega}(x) - \omega_{\mathbf{o}}(x) = \phi_*^{(2)}(\lambda'_{\mathbf{o}} g(x) + t_x(\hat{\lambda} - \lambda_{\mathbf{o}})' g(x))(\hat{\lambda} - \lambda_{\mathbf{o}})' g(x), \quad (49)$$

for each  $x \in \mathcal{X}$ .

First, consider the case (i) when  $\tilde{\zeta}_K \kappa_{\mathbf{o}, n} \lesssim 1$ . Hölder's inequality and Lemma 4(ii) imply

$$\sup_{x \in \mathcal{X}} |t_x(\hat{\lambda} - \lambda_{\mathbf{o}})' g(x)| \leq \|\hat{\lambda} - \lambda_{\mathbf{o}}\|_1 \tilde{\zeta}_K = O_p(\tilde{\zeta}_K \kappa_{\mathbf{o}, n}) = O_p(1). \quad (50)$$

The assumption  $\sup_{x \in \mathcal{X}} |\omega_{\mathbf{o}}(x) - \omega_0(x)| \lesssim 1$  and (50) imply  $\mathbb{P}\{\mathcal{E}_n\} \rightarrow 1$ , where  $\mathcal{E}_n$  is the event that  $\phi_*^{(2)}(\lambda'_{\mathbf{o}} g(x) + t_x(\hat{\lambda} - \lambda_{\mathbf{o}})' g(x))$  lies in a bounded set for all  $x \in \mathcal{X}$ . On the event  $\mathcal{E}_n$ , (49) and (50) imply

$$\begin{aligned} \mathbb{E}_n[\{\hat{\omega}(X) - \omega_{\mathbf{o}}(X)\}^2] &\lesssim (\hat{\lambda} - \lambda_{\mathbf{o}})' \mathbb{E}_n[g(X)g(X)'] (\hat{\lambda} - \lambda_{\mathbf{o}}) \\ &\leq \|\hat{\lambda} - \lambda_{\mathbf{o}}\|_1^2 \|\mathbb{E}_n[g(X)g(X)']\|_{\infty} \\ &= O_p(\kappa_{\mathbf{o}n}^2 \xi_n), \end{aligned}$$

where the second inequality follows from Hölder's inequality and the equality follows from Lemma 4(ii) and the definition of  $\xi_n$ .

Now consider the case (ii) when  $\phi_*^{(2)}$  is bounded from above and away from zero. In this case, it is easy to see that we still have  $\mathbb{E}_n[\{\hat{\omega}(X) - \omega_{\mathbf{o}}(X)\}^2] = O_p(\kappa_{\mathbf{o}n}^2 \xi_n)$  from (49).

Therefore for both cases, on the event  $\mathcal{E}_n$ , the triangle inequality, the result  $\mathbb{E}_n[\{\hat{\omega}(X) - \omega_{\mathbf{o}}(X)\}^2] = O_p(\kappa_{\mathbf{on}}^2 \xi_n)$ , and the assumption  $\sqrt{\mathbb{E}[\{\omega_{\mathbf{o}}(X) - \omega_0(X)\}^2]} \lesssim \mathbf{s}_{\mathbf{o},n}$  yield the conclusion in (23).

Proofs of  $\hat{\theta} \xrightarrow{P} \theta_0$  and (24) are similar to those of Theorem 1, and thus omitted.

### B.2 Proof of Theorem 4

We employ the notation in (47). By the Karush–Kuhn–Tucker (KKT) condition of  $\hat{\lambda}$  in (14) for the high-dimensional case, an expansion around  $\hat{\lambda} = \lambda_{\mathbf{o}}$  yields

$$0 = Q_n^{(1)}(\hat{\lambda}) + \alpha_n \hat{\kappa} = Q_n^{(1)}(\lambda_{\mathbf{o}}) + c_* \mathbb{E}_n[g(X)g(X)'](\hat{\lambda} - \lambda_{\mathbf{o}}) + \alpha_n \hat{\kappa},$$

where  $Q_n(\lambda) = \mathbb{E}_n[\phi_*(\lambda'g(X)) - \lambda'r(X)]$  and  $Q_n^{(1)}(\lambda) = \mathbb{E}_n[\phi_*^{(1)}(\lambda'g(X))g(X) - r(X)]$  is its first derivative. Since  $\omega_{\mathbf{o}}(\cdot) = \phi_*^{(1)}(\lambda'_{\mathbf{o}}g(\cdot))$ , an expansion of  $\frac{1}{n} \sum_{i=1}^n \phi_*^{(1)}(\hat{\lambda}'g_i)h_i$  around  $\hat{\lambda} = \lambda_{\mathbf{o}}$  yields

$$\hat{\theta}_{\text{DB}} = \frac{1}{n} \sum_{i=1}^n \omega_{\mathbf{o}i} h_i + \frac{1}{n} \sum_{i=1}^n c_* h_i g_i' \{(\hat{\lambda} - \lambda_{\mathbf{o}}) + \alpha_n \hat{\Theta} \hat{\kappa}\}.$$

By plugging in the form of  $\alpha_n \hat{\kappa}$  from the KKT condition to the above equation, we obtain

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n c_* h_i g_i' \{(\hat{\lambda} - \lambda_{\mathbf{o}}) + \alpha_n \hat{\Theta} \hat{\kappa}\} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n c_* h_i g_i' \{(\hat{\lambda} - \lambda_{\mathbf{o}}) - \hat{\Theta} [Q_n^{(1)}(\lambda_{\mathbf{o}}) + \mathbb{E}_n[g(X)g(X)'](\hat{\lambda} - \lambda_{\mathbf{o}})]\} \\ &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n c_* h_i g_i' \hat{\Theta} \mathbb{E}_n[\omega_{\mathbf{o}}(X)g(X) - r(X)] + T_{\Delta}, \end{aligned}$$

where  $T_{\Delta} = \frac{1}{\sqrt{n}} \sum_{i=1}^n c_* h_i g_i' (I - \mathbb{E}_n[g(X)g(X)']\hat{\Theta})(\hat{\lambda} - \lambda_{\mathbf{o}})$ . Combining these results and the definition of  $\hat{\beta}_{\text{DB}}$ , we obtain the following decomposition:

$$\sqrt{n}(\hat{\theta}_{\text{DB}} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{r_i^h - \theta_0 + \omega_{0i}(h_i - h_i^X)\} + T_1 + T_2 + T_3 + T_4 + T_5 + T_{\Delta},$$

where

$$T_1 = -c_* \frac{1}{\sqrt{n}} \sum_{i=1}^n [\hat{\beta}'_{\text{DB}}(\omega_{0i}g_i - r_i) - (\omega_{0i}\tilde{h}_i^X - \tilde{r}_i^h)],$$

$$T_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{\mathbf{o}i} - \omega_{0i})(\tilde{h}^X - \hat{\beta}'_{\text{DB}}g_i),$$

$$T_3 = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{\mathbf{o}i} - \omega_{0i})(h_i^X - \tilde{h}_i^X),$$

$$T_4 = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{\mathbf{0}i} - \omega_{0i})(h_i - h_i^X),$$

$$T_5 = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\omega_{0i}(h_i^X - \tilde{h}_i^X) + (\tilde{r}_i^h - r_i^h)].$$

Condition DB guarantees  $T_1 \xrightarrow{P} 0$ . By the Cauchy–Schwarz inequality,

$$|T_2| \leq \sqrt{n} \sqrt{\frac{1}{n} \sum_{i=1}^n (\omega_{\mathbf{0}i} - \omega_{0i})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (\tilde{h}^X - \hat{\beta}'_{\text{DB}} g_i)^2} \xrightarrow{P} 0,$$

where the equality follows from Chebychev’s inequality for the term  $\frac{1}{n} \sum_{i=1}^n (\omega_{\mathbf{0}i} - \omega_{0i})^2$  and Condition DB.

For  $T_3$ , the Cauchy–Schwarz inequality and the assumptions in the theorem imply  $\mathbb{E}[T_3] \lesssim \sqrt{n} s_n \tau_n \rightarrow 0$ . Also, Chebychev’s inequality implies  $T_3 - \mathbb{E}[T_3] \xrightarrow{P} 0$ . Combining these results, we obtain  $T_3 \xrightarrow{P} 0$ . Note that both  $T_4$  and  $T_5$  have zero mean. Thus, Chebyshev’s inequality implies  $T_4 = O_p(s_n) = o_p(1)$  and  $T_5 = O_p(\tau_n) = o_p(1)$ . Finally, by Hölder’s inequality, we have

$$T_\Delta \lesssim \sqrt{n} \left\| \frac{1}{n} \sum_{i=1}^n h_i g_i \right\|_\infty \|I - \mathbb{E}_n[g(X)g(X)'] \hat{\Theta}\|_1 \|\hat{\lambda} - \lambda_{\mathbf{0}}\|_1 = o_p(1),$$

under the assumptions of this theorem.

Combining these results, we obtain

$$\sqrt{n}(\hat{\theta}_{\text{DB}} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{r_i^h - \theta_0 + \omega_{0i}(h_i - h_i^X)\} + o_p(1),$$

and the conclusion follows by a central limit theorem.

### B.3 Proof of Theorem 5

First, we show  $|\hat{\Lambda} - \Lambda_*| = O_p(\gamma_n)$ , where  $\gamma_n = \sqrt{\zeta_{\mathbf{s}}^2/n}$ . Recall  $\hat{\Lambda} = \arg \max_{\Lambda \in \mathbb{R}^s} \hat{Q}_{\mathbf{s}}(\Lambda)$ , where

$$\hat{Q}_{\mathbf{s}}(\Lambda) = \mathbb{E}_n[\Lambda' r_{\mathbf{s}}(X) - \phi_*(\Lambda' g_{\mathbf{s}}(X))].$$

By Condition I',  $\hat{Q}_{\mathbf{s}}(\Lambda)$  is strictly concave in  $\Lambda$ . By taking the derivative, we have  $\hat{Q}_{\mathbf{s}}^{(1)}(\Lambda_*) = \mathbb{E}_n[r_{\mathbf{s}}(X) - \phi_*^{(1)}(\Lambda_*' g_{\mathbf{s}}(X))g_{\mathbf{s}}(X)]$ . Also note that  $\mathbb{E}[r_{\mathbf{s}}(X) - \phi_*^{(1)}(\Lambda_*' \times g_{\mathbf{s}}(X))g_{\mathbf{s}}(X)] = 0$  because  $\Lambda_*$  minimizes  $\mathbb{E}[\Lambda' r_{\mathbf{s}}(X) - \phi_*(\Lambda' g_{\mathbf{s}}(X))]$ . Thus, by Assumption S' and Chebyshev’s inequality, we have  $\hat{Q}_{\mathbf{s}}^{(1)}(\Lambda_*) = O_p(\sqrt{\zeta_{\mathbf{s}}^2/n})$ . The rest of the proof is similar to steps 2–4 in Lemma 3(iv), and thus is omitted.

Next, by an expansion of  $\tilde{\theta} = \frac{1}{n} \sum_{i=1}^n \phi_*^{(1)}(\hat{\Lambda}' g_{\mathbf{s}i}) h_i$  around  $\hat{\Lambda} = \Lambda_*$ , we obtain

$$\sqrt{n}(\tilde{\theta} - \theta_0 + b) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\Phi_i + v_{1i} + v_{2i} + v_{3i}) + T_1 + T_2 + T_3,$$

where

$$\begin{aligned}
 T_1 &= \mathbb{E}[\phi_*^{(2)}(\Lambda'_* g_{si}) h_i g_{si}]' \sqrt{n}(\hat{\Lambda} - \Lambda_*) + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{*i} \tilde{h}_i^X - \tilde{r}_i^h), \\
 T_2 &= \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_*^{(2)}(\Lambda'_* g_{si}) h_i g_{si} - \mathbb{E}[\phi_*^{(2)}(\Lambda'_* g_{si}) h_i g_{si}] \right)' (\hat{\Lambda} - \Lambda_*), \\
 T_3 &= \frac{1}{2}(\hat{\Lambda} - \Lambda_*)' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_*^{(3)}(\tilde{\Lambda}' g_{si}) h_i g_{si} g'_{si} \right) (\hat{\Lambda} - \Lambda_*),
 \end{aligned}$$

and  $\tilde{\Lambda}$  is on the line joining  $\hat{\Lambda}$  and  $\Lambda_*$ . By Condition I' and Chebyshev and Cauchy-Schwarz inequalities, we have

$$|T_2| \leq \sqrt{n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_*^{(2)}(\Lambda'_* g_{si}) h_i g_{si} - \mathbb{E}[\phi_*^{(2)}(\Lambda'_* g_{si}) h_i g_{si}] \right| |\hat{\Lambda} - \Lambda_*| = O_p(\zeta_s \gamma_n).$$

For  $T_3$ , similarly we have

$$|T_3| \leq \sqrt{n} |\hat{\Lambda} - \Lambda_*|^2 \left| \frac{1}{n} \sum_{i=1}^n \phi_*^{(3)}(\tilde{\Lambda}' g_{si}) h_i g_{si} g'_{si} \right|^2 = O_p(\sqrt{n} \zeta_s^2 \gamma_n^2).$$

We now consider  $T_1$ . By expanding the first-order condition of  $\hat{\Lambda}$ ,

$$\begin{aligned}
 0 &= \frac{1}{n} \sum_{i=1}^n \{ \phi_*^{(1)}(\hat{\Lambda}' g_{si}) g_{si} - r_{si} \} \\
 &= \frac{1}{n} \sum_{i=1}^n (\omega_{*i} g_{si} - r_{si}) + \frac{1}{n} \sum_{i=1}^n \phi_*^{(2)}(\bar{\Lambda}' g_{si}) g_{si} g'_{si} (\hat{\Lambda} - \Lambda_*),
 \end{aligned}$$

where  $\bar{\Lambda}$  lies on the line joining  $\hat{\Lambda}$  and  $\Lambda_*$ . Denote  $\Sigma_s = \mathbb{E}[\phi_*^{(2)}(\Lambda'_* g_{si}) g_{si} g'_{si}]$  and  $\bar{\Sigma}_s = \frac{1}{n} \sum_{i=1}^n \phi_*^{(2)}(\bar{\Lambda}' g_{si}) g_{si} g'_{si}$ . By solving the above equation for  $\hat{\Lambda} - \Lambda_*$  and inserting to  $T_1$ , we have

$$\begin{aligned}
 T_1 &= -\mathbb{E}[\phi_*^{(2)}(\Lambda'_* g_{si}) h_i g_{si}]' \bar{\Sigma}_s^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{*i} g_{si} - r_{si}) + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{*i} \tilde{h}_i^X - \tilde{r}_i^h) \\
 &= T_{11} + T_{12} + T_{13},
 \end{aligned}$$

where

$$\begin{aligned}
 T_{11} &= -\mathbb{E}[\phi_*^{(2)}(\Lambda'_* g_{si}) h_i g_{si}]' (\bar{\Sigma}_s^{-1} - \Sigma_s^{-1}) \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{*i} g_{si} - r_{si}), \\
 T_{12} &= -\mathbb{E}[\phi_*^{(2)}(\Lambda'_* g_{si}) h_i g_{si}]' \Sigma_s^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{0i} g_{si} - r_{si}) + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{0i} \tilde{h}_i^X - \tilde{r}_i^h),
 \end{aligned}$$

$$T_{13} = -\mathbb{E}[\phi_*^{(2)}(\Lambda'_* g_{\mathbf{s}i}) h_i g_{\mathbf{s}i}]' \Sigma_{\mathbf{s}}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{*i} - \omega_{0i}) g_{\mathbf{s}i} + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{*i} - \omega_{0i}) \tilde{h}_i^X.$$

For  $T_{11}$ , we apply a similar argument used to bound  $T_{11}$  in Theorem 2 but for iid data, which yields  $|T_{11}| = O_p(\sqrt{n} \zeta_{\mathbf{s}}^4 \gamma_n^2)$ . Note that  $\mathbb{E}[T_{12}] = 0$ . By Condition N'(2) and Chebyshev's inequality, we have  $T_{12} = o_p(1)$ . Also, the definition of  $\tilde{h}_i^X$  implies  $T_{13} = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{*i} - \omega_{0i}) (\tilde{h}_i^X - \beta'_{\mathbf{s}} g_{\mathbf{s}i}) = 0$ . Combining these results, we have

$$\sqrt{n}(\tilde{\theta} - \theta_0 + b) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\Phi_i + v_{1i} + v_{2i} + v_{3i}) + r_n,$$

where  $r_n = O_p(\zeta_{\mathbf{s}}^6 / \sqrt{n}) = o_p(1)$  under the assumptions in this theorem. The conclusion follows by applying a central limit theorem for iid data.

#### B.4 Proof of Theorem 6

Recall  $\omega_{\mathbf{s}}(x) = \phi_*^{(1)}(\lambda'_{\mathbf{o}\mathbf{s}} g_{\mathbf{o}\mathbf{s}}(x))$ . By an expansion of the debiased estimator,

$$\hat{\theta}_{\text{TD}} = \frac{1}{n} \sum_{i=1}^n \phi_*^{(1)}(\hat{\lambda}'_{\text{TD}} g_i) h_i = \frac{1}{n} \sum_{i=1}^n \phi_*^{(1)}(\hat{\Lambda}'_{\mathbf{s}} g_{\mathbf{s}i}) h_i$$

around  $\hat{\Lambda}_{\mathbf{s}} = \lambda_{\mathbf{o}\mathbf{s}}$ , we obtain

$$\sqrt{n}(\hat{\theta}_{\text{TD}} - \theta_0 + \tilde{b}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\Phi_i + \tilde{v}_{1i} + \tilde{v}_{2i} + \tilde{v}_{3i}) + T_1 + T_2 + T_3,$$

where

$$\begin{aligned} T_1 &= \sqrt{n} \mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{o}\mathbf{s}} g_{\mathbf{s}i}) h_i g_{\mathbf{s}i}]' (\hat{\Lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}}) + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{\mathbf{s}i} \tilde{h}_{\text{TD}i}^X - \tilde{r}_{\text{TD}i}^h), \\ T_2 &= \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \{ \phi_*^{(2)}(\lambda'_{\mathbf{o}\mathbf{s}} g_{\mathbf{s}i}) h_i g_{\mathbf{s}i} - \mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{o}\mathbf{s}} g_{\mathbf{s}i}) h_i g_{\mathbf{s}i}] \} \right]' (\hat{\Lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}}), \\ T_3 &= \frac{1}{2} (\hat{\Lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}})' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_*^{(3)}(\tilde{\Lambda}'_{\mathbf{s}} g_{\mathbf{s}i}) h_i g_{\mathbf{s}i} g'_{\mathbf{s}i} \right) (\hat{\Lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}}), \end{aligned}$$

and  $\tilde{\Lambda}_{\mathbf{s}}$  is on the line joining  $\hat{\Lambda}_{\mathbf{s}}$  and  $\lambda_{\mathbf{o}\mathbf{s}}$ . Since Condition TD(3) implies  $\mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{o}\mathbf{s}} g_{\mathbf{s}}) h]^2 = O(1)$ , Chebyshev's inequality yields

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \{ \phi_*^{(2)}(\lambda'_{\mathbf{o}\mathbf{s}} g_{\mathbf{s}i}) h_i g_{\mathbf{s}i} - \mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{o}\mathbf{s}} g_{\mathbf{s}i}) h_i g_{\mathbf{s}i}] \} \right| = O_p(\sqrt{\zeta_{\mathbf{s}}^2/n}).$$

Thus, by the Cauchy–Schwarz inequality and Lemma 5(ii), it follows

$$|T_2| \leq \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n \{ \phi_*^{(2)}(\lambda'_{\mathbf{o}\mathbf{s}} g_{\mathbf{s}i}) h_i g_{\mathbf{s}i} - \mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{o}\mathbf{s}} g_{\mathbf{s}i}) h_i g_{\mathbf{s}i}] \} \right| |\hat{\Lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}}| = O_p(\zeta_{\mathbf{s}} \tilde{\gamma}_n).$$

For  $T_3$ , note that

$$|T_3| \leq \sqrt{n} |g_{\mathbf{s}}|^2 |\hat{\Lambda}_{\mathbf{s}} - \lambda_{\mathbf{os}}|^2 \sqrt{\frac{1}{n} \sum_{i=1}^n \phi_*^{(3)}(\tilde{\Lambda}'_{\mathbf{s}} g_{\mathbf{s}i})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n h_i^2} = O_p(\sqrt{n} \zeta_{\mathbf{s}}^2 \tilde{\gamma}_n^2),$$

where the first inequality follows from Cauchy–Schwarz inequality, and the equality follows from the law of large numbers, Condition TD(3), and Lemma 5(ii).

Now we consider  $T_1$ . By Lemma 5(i), we have

$$\hat{\Lambda}_{\mathbf{s}} - \lambda_{\mathbf{os}} = -\hat{\Theta}_{\mathbf{s}} \frac{1}{n} \sum_{i=1}^n (\omega_{\mathbf{s}i} g_{\mathbf{s}i} - r_{\mathbf{s}i}) + \tilde{\Delta},$$

where  $\tilde{\Delta} = (I_{\mathbf{s}} - \hat{\Theta}_{\mathbf{s}} Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}}))(\hat{\Lambda}_{\mathbf{s}} - \lambda_{\mathbf{os}})$  and  $Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}}) = \mathbb{E}_n[\phi_*^{(2)}(\bar{\lambda}'_{\mathbf{s}} g_{\mathbf{s}}) g_{\mathbf{s}} g'_{\mathbf{s}}]$ . Also let  $Q^{(2)}(\lambda_{\mathbf{os}}) = \mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{os}} g_{\mathbf{s}}) g_{\mathbf{s}} g'_{\mathbf{s}}]$ . Note that  $T_1$  is decomposed as  $T_1 = T_{11} + \dots + T_{14}$ , where

$$\begin{aligned} T_{11} &= -\sqrt{n} \mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{os}} g_{\mathbf{s}i}) h_i g_{\mathbf{s}i}]' Q^{(2)}(\lambda_{\mathbf{os}})^{-1} \frac{1}{n} \sum_{i=1}^n (\omega_{0i} g_{\mathbf{s}i} - r_{\mathbf{s}i}) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{0i} \tilde{h}_{\text{TD}i}^X - \tilde{r}_{\text{TD}i}^h), \\ T_{12} &= -\sqrt{n} \mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{os}} g_{\mathbf{s}i}) h_i g_{\mathbf{s}i}]' Q^{(2)}(\lambda_{\mathbf{os}})^{-1} \frac{1}{n} \sum_{i=1}^n (\omega_{\mathbf{s}i} - \omega_{0i}) g_{\mathbf{s}i} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{\mathbf{s}i} - \omega_{0i}) \tilde{h}_i^X, \\ T_{13} &= -\sqrt{n} \mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{os}} g_{\mathbf{s}i}) h_i g_{\mathbf{s}i}]' (\hat{\Theta} - Q^{(2)}(\lambda_{\mathbf{os}})^{-1}) \frac{1}{n} \sum_{i=1}^n (\omega_{\mathbf{s}i} g_{\mathbf{s}i} - r_{\mathbf{s}i}), \\ T_{14} &= \sqrt{n} \mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{os}} g_{\mathbf{s}i}) h_i g_{\mathbf{s}i}] \tilde{\Delta}. \end{aligned}$$

For  $T_{11}$ , Condition TD and Chebychev's inequality imply

$$T_{11} = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \{\tilde{\beta}'_{\mathbf{s}} (\omega_{0i} g_{\mathbf{s}i} - r_{\mathbf{s}i}) - (\omega_{0i} \tilde{h}_i^X - \tilde{r}_i^h)\} \xrightarrow{P} 0.$$

By the definition, we have  $T_{12} = -\frac{1}{\sqrt{n}} \sum_{i=1}^n (\omega_{\mathbf{s}i} - \omega_{0i})(\tilde{\beta}'_{\mathbf{s}} g_{\mathbf{s}i} - \tilde{h}_i^X) = 0$ . To bound  $T_{13}$ , note that  $\mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{os}} g_{\mathbf{s}i}) h_i g_{\mathbf{s}i}] = O_p(\zeta_{\mathbf{s}})$ . By the Cauchy–Schwarz inequality, Lemma 5(iv), and Condition TD(2), we have

$$\begin{aligned} |T_{13}| &= \left| \sqrt{n} \mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{os}} g_{\mathbf{s}i}) h_i^X g_{\mathbf{s}i}]' (\hat{\Theta} - Q^{(2)}(\lambda_{\mathbf{os}})^{-1}) \frac{1}{n} \sum_{i=1}^n (\omega_{\mathbf{s}i} g_{\mathbf{s}i} - r_{\mathbf{s}i}) \right| \\ &= O_p(\sqrt{n} \zeta_{\mathbf{s}} Q_n \tilde{\gamma}_n). \end{aligned}$$

Similarly, by the Cauchy–Schwarz inequality, Lemma 5(ii) and (v), and the relation between  $\ell_1$ - and  $\ell_2$ -norms, it holds

$$\begin{aligned} |T_{14}| &= |\sqrt{n}\mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{o}\mathbf{s}}g_{\mathbf{s}i})h_i^X g_{\mathbf{s}i}]'(I_{\mathbf{s}} - \hat{\Theta}Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}}))(\hat{\lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}})| \\ &\leq \sqrt{n}\mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{o}\mathbf{s}}g_{\mathbf{s}i})h_i^X g_{\mathbf{s}i}]'|I_{\mathbf{s}} - \hat{\Theta}Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}})|\|\hat{\lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}}\|_1 \\ &= O_p(\sqrt{n}\kappa_{\mathbf{o},n}^2\zeta_{\mathbf{s}}^4 + \sqrt{n}\zeta_{\mathbf{s}}\kappa_{\mathbf{o},n}\varrho_n). \end{aligned}$$

Combining these results, we obtain

$$\sqrt{n}(\hat{\theta}_{\text{TD}} - \theta_0 + \tilde{b}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\Phi_i + \tilde{v}_{1i} + \tilde{v}_{2i} + \tilde{v}_{3i}) + r_n,$$

where  $r_n = O_p(\sqrt{n}\kappa_{\mathbf{o},n}^2\zeta_{\mathbf{s}}^4 + \sqrt{n}\tilde{\gamma}_n\zeta_{\mathbf{s}}\varrho_n + \sqrt{n}\zeta_{\mathbf{s}}^2\tilde{\gamma}_n^2) = o_p(1)$  under the assumptions of this theorem. The conclusion follows by applying a central limit theorem.

### B.5 Lemmas

LEMMA 4. *Under the conditions of Theorem 3, it holds*

- (i)  $\Pr\{\frac{1}{2}\mathcal{E}(\hat{\lambda}) + \alpha_n\|\hat{\lambda} - \lambda_{\mathbf{o}}\|_1 \leq 4\mathcal{E}(\lambda_{\mathbf{o}}) + \frac{16\alpha_n^2s}{\phi_{\delta_{\lambda_{\mathbf{o}}}^2}^{\varrho}}\} \geq 1 - \varepsilon,$
- (ii)  $\mathcal{E}(\hat{\lambda}) = O_p(\kappa_{\mathbf{o}n}\sqrt{\log K/n})$  and  $\|\hat{\lambda} - \lambda_{\mathbf{o}}\|_1 = O_p(\kappa_{\mathbf{o}n}).$

LEMMA 5. *Let  $Q(\lambda_{\mathbf{s}}) = \mathbb{E}[\phi_*(\lambda'_{\mathbf{s}}g_{\mathbf{s}}) - \lambda'_{\mathbf{s}}r_{\mathbf{s}}]$  and  $Q_n(\lambda_{\mathbf{s}}) = \mathbb{E}_n[\phi_*(\lambda'_{\mathbf{s}}g_{\mathbf{s}}) - \lambda'_{\mathbf{s}}r_{\mathbf{s}}]$ . Under the conditions of Theorem 6, it holds*

- (i)  $\hat{\Lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}} = -\hat{\Theta}\frac{1}{n}\sum_{i=1}^n(\omega_{\mathbf{s}i}g_{\mathbf{s}i} - r_{\mathbf{s}i}) + \tilde{\Delta}$ , where  $\tilde{\Delta} = (I_{\mathbf{s}} - \hat{\Theta}_{\mathbf{s}}Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}}))(\hat{\lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}})$ , and  $\bar{\lambda}_{\mathbf{s}}$  is on the line between  $\hat{\lambda}_{\mathbf{s}}$  and  $\lambda_{\mathbf{o}\mathbf{s}}$ ,
- (ii)  $|\hat{\Lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}}| = O_p(\tilde{\gamma}_n)$ , where  $\tilde{\gamma}_n = \kappa_{\mathbf{o},n} \vee \sqrt{\mathbf{s}\log K/n}$ ,
- (iii)  $|Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}}) - Q^{(2)}(\lambda_{\mathbf{o}\mathbf{s}})| = O_p(\kappa_{\mathbf{o},n}\zeta_{\mathbf{s}}^3)$ ,
- (iv)  $|\frac{1}{n}\sum_{i=1}^n(\omega_{\mathbf{s}i}g_{\mathbf{s}i} - r_{\mathbf{s}i})| = O_p(\tilde{\gamma}_n)$ ,
- (v)  $|I_{\mathbf{s}} - \hat{\Theta}_{\mathbf{s}}Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}})| = O_p(\kappa_{\mathbf{o},n}\zeta_{\mathbf{s}}^3 + \varrho_n).$

#### *Proof of Lemma 4(i)*

Pick any  $\varepsilon > 0$  small enough and  $n \in \mathbb{N}$  large enough to satisfy Condition H. Then set  $M = \frac{Q_{\mathbf{o}}}{2\sigma_{\varepsilon,n}}$  and take  $\bar{\lambda} = t\hat{\lambda} + (1-t)\lambda_{\mathbf{o}}$  with  $t = \frac{M}{M + \|\hat{\lambda} - \lambda_{\mathbf{o}}\|_1}$ . Due to the definition of  $\hat{\lambda}$  in (14) and convexity of its objective function, we have

$$\begin{aligned} &\mathbb{E}_n[\phi_*(\bar{\lambda}'g(X)) - \bar{\lambda}'r(X)] + \alpha_n\|\bar{\lambda}\|_1 \\ &\leq \mathbb{E}_n[\phi_*(\lambda'_{\mathbf{o}}g(X)) - \lambda'_{\mathbf{o}}r(X)] + \alpha_n\|\lambda_{\mathbf{o}}\|_1, \end{aligned}$$

and thus

$$\begin{aligned} \mathcal{E}(\bar{\lambda}) + \alpha_n \|\bar{\lambda}\|_1 &\leq -\{v_n(\bar{\lambda}) - v_n(\lambda_{\mathbf{o}})\} + \mathcal{E}(\lambda_{\mathbf{o}}) + \alpha_n \|\lambda_{\mathbf{o}}\|_1 \\ &\leq \mathcal{E}(\lambda_{\mathbf{o}}) + \alpha_n \|\lambda_{\mathbf{o}}\|_1 + \frac{Q_{\mathbf{o}}}{2}, \end{aligned} \quad (51)$$

with probability at least  $1 - \varepsilon$ , where the second inequality follows from Condition H(1) combined with  $\|\bar{\lambda} - \lambda_{\mathbf{o}}\|_1 = \frac{M\|\hat{\lambda} - \lambda_{\mathbf{o}}\|_1}{M + \|\hat{\lambda} - \lambda_{\mathbf{o}}\|_1} \leq M$ . Hereafter, all inequalities involving  $\bar{\lambda}$  hold true with probability at least  $1 - \varepsilon$ .

Note that  $\lambda = \lambda_{S_{\lambda_{\mathbf{o}}}} + \lambda_{S_{\lambda_{\mathbf{o}}}^c}$ ,  $\lambda_{\mathbf{o}, S_{\lambda_{\mathbf{o}}}} = \lambda_{\mathbf{o}}$ , and  $\lambda_{\mathbf{o}, S_{\lambda_{\mathbf{o}}}^c} = 0$ . Thus, (51) and the triangle inequality imply

$$\begin{aligned} \mathcal{E}(\bar{\lambda}) + \alpha_n \|\bar{\lambda}_{S_{\lambda_{\mathbf{o}}}^c}\|_1 &\leq \mathcal{E}(\lambda_{\mathbf{o}}) + \alpha_n \|\bar{\lambda}_{S_{\lambda_{\mathbf{o}}}} - \lambda_{\mathbf{o}}\|_1 + \frac{Q_{\mathbf{o}}}{2} \\ &\leq Q_{\mathbf{o}} + \alpha_n \|\bar{\lambda}_{S_{\lambda_{\mathbf{o}}}} - \lambda_{\mathbf{o}}\|_1, \end{aligned} \quad (52)$$

where the second inequality follows from  $\mathcal{E}(\lambda_{\mathbf{o}}) \leq \frac{Q_{\mathbf{o}}}{2}$  (due to the definition of  $Q_{\mathbf{o}}$ ). Thus, the triangle inequality yields

$$\mathcal{E}(\bar{\lambda}) + \alpha_n \|\bar{\lambda} - \lambda_{\mathbf{o}}\|_1 \leq Q_{\mathbf{o}} + 2\alpha_n \|\bar{\lambda}_{S_{\lambda_{\mathbf{o}}}} - \lambda_{\mathbf{o}}\|_1. \quad (53)$$

In order to bound the right-hand side of (53), we consider two cases: (I)  $2\alpha_n \|\bar{\lambda}_{S_{\lambda_{\mathbf{o}}}} - \lambda_{\mathbf{o}}\|_1 < Q_{\mathbf{o}}$ , and (II)  $2\alpha_n \|\bar{\lambda}_{S_{\lambda_{\mathbf{o}}}} - \lambda_{\mathbf{o}}\|_1 \geq Q_{\mathbf{o}}$ .

**Case (I)**  $2\alpha_n \|\bar{\lambda}_{S_{\lambda_{\mathbf{o}}}} - \lambda_{\mathbf{o}}\|_1 < Q_{\mathbf{o}}$ .

In this case, (53) and Condition H(3) imply

$$\mathcal{E}(\bar{\lambda}) + \alpha_n \|\bar{\lambda} - \lambda_{\mathbf{o}}\|_1 < 2Q_{\mathbf{o}} \leq \frac{\alpha_n M}{2}, \quad (54)$$

and thus  $\|\bar{\lambda} - \lambda_{\mathbf{o}}\|_1 \leq \frac{M}{2}$ .

**Case (II)**  $2\alpha_n \|\bar{\lambda}_{S_{\lambda_{\mathbf{o}}}} - \lambda_{\mathbf{o}}\|_1 \geq Q_{\mathbf{o}}$ .

In this case, (52) and  $\lambda_{\mathbf{o}, S_{\lambda_{\mathbf{o}}}^c} = 0$  guarantees

$$\begin{aligned} \|\bar{\lambda}_{S_{\lambda_{\mathbf{o}}}^c} - \lambda_{\mathbf{o}, S_{\lambda_{\mathbf{o}}}^c}\|_1 &= \|\bar{\lambda}_{S_{\lambda_{\mathbf{o}}}^c}\|_1 \\ &\leq 3\|\bar{\lambda}_{S_{\lambda_{\mathbf{o}}}} - \lambda_{\mathbf{o}, S_{\lambda_{\mathbf{o}}}}\|_1 \\ &\leq \frac{3\sqrt{s}}{\phi_{S_{\lambda_{\mathbf{o}}}}} \sqrt{(\bar{\lambda} - \lambda_{\mathbf{o}})' \mathbb{E}[g(X)g(X)'](\bar{\lambda} - \lambda_{\mathbf{o}})}, \end{aligned}$$

where the last inequality follows from Condition C. Observe that

$$\mathcal{E}(\bar{\lambda}) + \alpha_n \|\bar{\lambda} - \lambda_{\mathbf{o}}\|_1 \leq 4\alpha_n \|\bar{\lambda}_{S_{\lambda_{\mathbf{o}}}} - \lambda_{\mathbf{o}}\|_1 \leq \frac{4\alpha_n \sqrt{s}}{\phi_{S_{\lambda_{\mathbf{o}}}}} \sqrt{(\bar{\lambda} - \lambda_{\mathbf{o}})' \mathbb{E}[g(X)g(X)'](\bar{\lambda} - \lambda_{\mathbf{o}})},$$

where the first inequality follows from (53) and the condition of Case (II), and the second inequality follows from (55) (note  $\lambda_{\mathbf{o}} = \lambda_{\mathbf{o}, S_{\lambda_{\mathbf{o}}}$ ). Now by using  $xy \leq x^2 + \frac{y^2}{4}$  for any  $x, y \in \mathbb{R}$ ,



we obtain

$$\begin{aligned}
& \frac{4\alpha_n\sqrt{s}}{\phi_{S_{\lambda_{\mathbf{o}}}}} \sqrt{(\bar{\lambda} - \lambda_{\mathbf{o}})' \mathbb{E}[g(X)g(X)'] (\bar{\lambda} - \lambda_{\mathbf{o}})} \\
& \leq \frac{1}{2} \left( \varrho(\bar{\lambda} - \lambda_{\mathbf{o}})' \mathbb{E}[g(X)g(X)'] (\bar{\lambda} - \lambda_{\mathbf{o}}) + \frac{16\alpha_n s}{\phi_{S_{\lambda_{\mathbf{o}}}^2} \varrho} \right) \\
& \leq \frac{1}{2} \left( \mathcal{E}(\bar{\lambda}) + \frac{16\alpha_n s}{\phi_{S_{\lambda_{\mathbf{o}}}^2} \varrho} \right),
\end{aligned}$$

where the second inequity follows from Condition H(2). Combining these results with the definition of  $Q_{\mathbf{o}}$ ,

$$\mathcal{E}(\bar{\lambda}) + \alpha_n \|\bar{\lambda} - \lambda_{\mathbf{o}}\|_1 \leq \frac{1}{2} \mathcal{E}(\bar{\lambda}) + \frac{8\alpha_n^2 s}{\phi_{S_{\lambda_{\mathbf{o}}}^2} \varrho} \leq \frac{1}{2} \mathcal{E}(\bar{\lambda}) + Q_{\mathbf{o}}, \quad (55)$$

which implies (by Condition H(3))  $\|\bar{\lambda} - \lambda_{\mathbf{o}}\|_1 \leq \frac{2\sigma_{\varepsilon} M}{\alpha_n} \leq \frac{M}{4}$ .

Therefore, for both cases, it holds  $\|\bar{\lambda} - \lambda_{\mathbf{o}}\|_1 \leq \frac{M}{2}$  and also  $\|\hat{\lambda} - \lambda_{\mathbf{o}}\|_1 \leq M$ , that is,  $\hat{\lambda}$  is close enough to  $\lambda_{\mathbf{o}}$  to invoke Condition H(1).

Repeat the proof above by replacing  $\bar{\lambda}$  with  $\hat{\lambda}$ . Then we obtain the counterparts of (54) and (55) with replacements of  $\bar{\lambda}$  with  $\hat{\lambda}$ , that is,

$$\frac{1}{2} \mathcal{E}(\hat{\lambda}) + \alpha_n \|\hat{\lambda} - \lambda_{\mathbf{o}}\|_1 \leq 2Q_{\mathbf{o}},$$

with probability at least  $1 - \varepsilon$ . Therefore, the conclusion follows.

#### *Proof of Lemma 4(ii)*

By setting  $\alpha_n \propto \sqrt{\frac{\log K}{n}}$ , Part (i) of this lemma implies

$$\frac{1}{2} \mathcal{E}(\hat{\lambda}) + \sqrt{\frac{\log K}{n}} \|\hat{\lambda} - \lambda_{\mathbf{o}}\|_1 = O_p \left( \mathcal{E}(\lambda_{\mathbf{o}}) \vee \frac{s \log K}{n} \right),$$

and the conclusion follows.

#### *Proof of Lemma 5(i)*

By the KKT conditions for  $\hat{\lambda}_{\mathbf{s}}$ , an expansion around  $\lambda_{\mathbf{o}\mathbf{s}}$  yields

$$\mathbf{0}_{\mathbf{s}} = \frac{1}{n} \sum_{i=1}^n (\omega_{\mathbf{s}i} g_{\mathbf{s}i} - r_{\mathbf{s}i}) + \alpha_n \hat{\kappa}_{\mathbf{s}} = \frac{1}{n} \sum_{i=1}^n (\omega_{\mathbf{s}i} g_{\mathbf{s}i} - r_{\mathbf{s}i}) + Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}}) (\hat{\lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}}) + \alpha_n \hat{\kappa}_{\mathbf{s}}, \quad (56)$$

where  $\bar{\lambda}_{\mathbf{s}}$  is on the line between  $\hat{\lambda}_{\mathbf{s}}$  and  $\lambda_{\mathbf{o}\mathbf{s}}$ . Thus, we have

$$\begin{aligned}
\hat{\Lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}} &= \hat{\lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}} + \hat{\Theta}_{\mathbf{s}} \alpha_n \hat{\kappa}_{\mathbf{s}} \\
&= \hat{\lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}} - \hat{\Theta}_{\mathbf{s}} \left[ \frac{1}{n} \sum_{i=1}^n (\omega_{\mathbf{s}i} g_{\mathbf{s}i} - r_{\mathbf{s}i}) + Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}}) (\hat{\lambda}_{\mathbf{s}} - \lambda_{\mathbf{o}\mathbf{s}}) \right],
\end{aligned}$$

where  $I_{\mathbf{s}}$  is an  $\mathbf{s} \times \mathbf{s}$  identity matrix, the first equality follows from the definition of  $\hat{\Lambda}_{\mathbf{s}}$ , and the second equality follows from (56). The conclusion follows by the definition of  $\tilde{\Delta}$ .

*Proof of Lemma 5(ii)*

By the definition of  $\hat{\Lambda}_{\mathbf{s}}$ ,

$$\begin{aligned} |\hat{\Lambda}_{\mathbf{s}} - \lambda_{\mathbf{os}}| &\leq |\hat{\lambda}_{\mathbf{s}} - \lambda_{\mathbf{os}}| + |\hat{\Theta}_{\mathbf{s}} \alpha_n \hat{\kappa}_{\mathbf{s}}| \\ &\leq \|\hat{\lambda}_{\mathbf{s}} - \lambda_{\mathbf{os}}\|_1 + |\hat{\Theta}_{\mathbf{s}} \alpha_n \hat{\kappa}_{\mathbf{s}}| \\ &\lesssim \kappa_{\mathbf{o},n} + \sqrt{\frac{\mathbf{s} \log K}{n}} \\ &= O_p\left(\kappa_{\mathbf{o},n} \vee \sqrt{\frac{\mathbf{s} \log K}{n}}\right), \end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from the relationship between the  $\ell_1$ - and  $\ell_2$ -norms, and the third inequality follows from Lemma 4(ii) and the assumption  $|\hat{\Theta}_{\mathbf{s}}| = O_p(1)$ .

*Proof of Lemma 5(iii)*

Note that

$$Q_n^{(2)}(\lambda_{\mathbf{os}}) = \mathbb{E}[\phi_*^{(2)}(\lambda'_{\mathbf{os}} g_{\mathbf{s}}) g_{\mathbf{s}} g'_{\mathbf{s}}], \quad Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}}) = \mathbb{E}_n[\phi_*^{(2)}(\bar{\lambda}'_{\mathbf{s}} g_{\mathbf{s}}) g_{\mathbf{s}} g'_{\mathbf{s}}],$$

and further denote  $Q_n^{(2)}(\lambda_{\mathbf{os}}) = \mathbb{E}_n[\phi_*^{(2)}(\lambda'_{\mathbf{os}} g_{\mathbf{s}}) g_{\mathbf{s}} g'_{\mathbf{s}}]$ . By Lemma 5(ii) and Condition TD(3), we have

$$\begin{aligned} &|Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}}) - Q_n^{(2)}(\lambda_{\mathbf{os}})| \\ &= |\mathbb{E}_n[\{\phi_*^{(2)}(\lambda'_{\mathbf{os}} g_{\mathbf{s}}) - \phi_*^{(2)}(\bar{\lambda}'_{\mathbf{s}} g_{\mathbf{s}})\} g_{\mathbf{s}} g'_{\mathbf{s}}]| \\ &\leq \zeta_{\mathbf{s}}^2 \left\{ \sup_{\Lambda: \|\Lambda - \lambda_{\mathbf{os}}\|_1 \lesssim \tilde{\gamma}_n} \frac{1}{n} \sum_{i=1}^n \phi_*^{(3)}(\lambda'_{\mathbf{s}} g_{\mathbf{s}i})^2 \right\}^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^n \{(\bar{\lambda}_{\mathbf{s}} - \lambda_{\mathbf{os}})' g_{\mathbf{s}}\}^2 \right\}^{1/2} \\ &= O_p(\kappa_{\mathbf{o},n} \zeta_{\mathbf{s}}^3). \end{aligned}$$

Thus, the triangle inequality and Lemma 3(i) imply

$$\begin{aligned} |Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}}) - Q_n^{(2)}(\lambda_{\mathbf{os}})| &\leq |Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}}) - Q_n^{(2)}(\lambda_{\mathbf{os}})| + |Q_n^{(2)}(\lambda_{\mathbf{os}}) - Q_n^{(2)}(\lambda_{\mathbf{os}})| \\ &= O_p(\kappa_{\mathbf{o},n} \zeta_{\mathbf{s}}^3) + O_p\left(\sqrt{\frac{\zeta_{\mathbf{s}}^2 \log \mathbf{s}}{n}}\right) \\ &= O_p(\kappa_{\mathbf{o},n} \zeta_{\mathbf{s}}^3). \end{aligned}$$

*Proof of Lemma 5(iv)*

By (56), we have

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n (\omega_{si} g_{si} - r_{si}) \right| &\leq |Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}})(\hat{\lambda}_{\mathbf{s}} - \lambda_{\mathbf{os}})| + |\alpha_n \hat{\kappa}_{\mathbf{s}}| \leq |Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}})| \|\hat{\lambda}_{\mathbf{s}} - \lambda_{\mathbf{os}}\|_1 + |\alpha_n \hat{\kappa}_{\mathbf{s}}| \\ &\lesssim \|\hat{\lambda}_{\mathbf{s}} - \lambda_{\mathbf{os}}\|_1 + |\alpha_n \hat{\kappa}_{\mathbf{s}}| = O_p\left(\kappa_{\mathbf{o},n} \vee \sqrt{\frac{\mathbf{s} \log K}{n}}\right), \end{aligned}$$

where the second inequality follows from the definition of the matrix norm  $|\cdot|$  and the relationship between the  $\ell_1$ - and  $\ell_2$ -norms, and the third inequality uses Lemma 4(iii) and Condition TD.

*Proof of Lemma 5(v)*

By triangle inequality, we have

$$|I_{\mathbf{s}} - \hat{\Theta}_{\mathbf{s}} Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}})| \leq \left| \{Q^{(2)}(\lambda_{\mathbf{os}})^{-1} - \hat{\Theta}_{\mathbf{s}}\} Q^{(2)}(\lambda_{\mathbf{os}}) \right| + \left| \hat{\Theta}_{\mathbf{s}} \{Q^{(2)}(\lambda_{\mathbf{os}}) - Q_n^{(2)}(\bar{\lambda}_{\mathbf{s}})\} \right|.$$

Condition TD guarantees  $Q^{(2)}(\lambda_{\mathbf{os}}) = O(1)$  and  $\hat{\Theta}_{\mathbf{s}} = O_p(1)$ . Thus, the conclusion follows by Lemma 5(iii).

APPENDIX C: ADDITIONAL TABLES

TABLE 5. Cross-sectional regression for other low-dimensional portfolios.

	Intercept	$\lambda_{\text{SDF}}$	$\lambda_{\text{RM}}$	$\lambda_{\text{SMB}}$	$\lambda_{\text{HML}}$	Adjusted $R^2$
Panel A: 10 momentum						
KL: No penalty	0.752 (21.715)	-0.168 (-10.056)				0.918
KL: $\alpha_n = 0.05$	0.716 (18.714)	-0.129 (-9.493)				0.908
3 Factors	2.365 (1.576)		-1.198 (-0.754)	-0.068 (-0.057)	-1.485 (-1.615)	0.815
Panel B: 25 long term reversal and size						
KL: No penalty	0.741 (8.023)	-0.215 (-5.049)				0.505
KL: $\alpha_n = 0.05$	0.382 (4.372)	-0.180 (-9.416)				0.785
3 Factors	0.702 (2.541)		0.219 (0.833)	0.111 (1.678)	0.633 (5.051)	0.754

*Note:* The estimated SDF is derived in a rolling window out-of-sample fashion from July 1963 to December 2010. Panel A presents results using 10 momentum portfolios, and Panel B is concerned with results using 25 long term reversal and size portfolios. The second column is the estimated constant in each model, the last column records the adjusted  $R^2$ , and the other columns summarize estimated price of risk. Numbers in the bracket are the corresponding t-values. In each panel, the first row is about the estimated SDF with KL when no penalty is imposed, the second row is the estimated SDF with KL when penalty level is at 0.05, and the third row is the seminal Fama-French three-factor models.

TABLE 6. Cross-sectional regression for intermediate dimensional portfolios.

	Intercept <sub>m</sub>	$\lambda_{\text{SDF}}$	$\lambda_{\text{RM}}$	$\lambda_{\text{SMB}}$	$\lambda_{\text{HML}}$	Adjusted $R^2$
Panel A: 100 size and book-to-market						
KL: No penalty	1.033 (52.744)	-0.926 (-11.532)				0.581
KL: $\alpha_n = 0.1$	0.725 (20.435)	-0.273 (-13.367)				0.652
3 Factors	1.575 (8.618)		-0.639 (-3.670)	0.190 (5.577)	0.439 (11.175)	0.627
Panel B: 49 industry						
KL: No penalty	0.800 (16.239)	-0.129 (-4.852)				0.329
KL: $\alpha_n = 0.1$	0.686 (0.686)	-0.065 (-0.065)				0.294
3 Factors	1.064 (6.229)		-0.008 (-0.047)	-0.096 (-0.923)	-0.109 (-1.151)	-0.002
Panel C: 25 long term reversal+25 short term reversal+25 momentum						
KL: No penalty	1.083 (48.960)	-1.919 (-10.698)				0.605
KL: $\alpha_n = 0.1$	1.130 (43.162)	-0.484 (-7.705)				0.441
3 Factors	1.416 (4.489)		-0.432 (-1.454)	0.293 (3.370)	0.012 (0.064)	0.153

*Note:* Cross-sectional regression results in the intermediate case. The estimated SDF is derived in a rolling window out-of-sample fashion from July 1963 to December 2010, using portfolios in each corresponding panel. Panel A presents results using 100 size and book-to-market portfolios, Panel B presents results using 49 industry portfolios, and Panel C presents results using 75 portfolios listed in the beginning of the panel. The second column is the estimated constant in each model, the last column records the adjusted  $R^2$ , and the other columns summarize estimated price of risk. Numbers in the bracket are the corresponding t-values. In each panel, the first row is about the estimated SDF with KL when no penalty is imposed, the second row is the estimated SDF with KL when penalty level is at 0.1, and the third row is the seminal Fama–French three-factor models.

## REFERENCES

- Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015), “Some new asymptotic theory for least squares series: Pointwise and uniform results.” *Journal of Econometrics*, 186, 345–366. [3]
- Bradley, R. C. (1985), “On the central limit question under absolute regularity.” *Annals of Probability*, 13, 1314–1325. [8]
- Chen, X. and T. Christensen (2015), “Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions.” *Journal of Econometrics*, 188, 447–465. [3]
- Hall, P. and C. C. Heyde (2014), *Martingale Limit Theory and Its Application*. Academic Press. [2]

Newey, W. K. and D. L. McFadden (1994), “Large sample estimation and hypothesis testing.” In *Handbook of Econometrics*, Vol. 4 (R. F. Engle and D. L. McFadden, eds.). Elsevier. [4]

---

Co-editor Andres Santos handled this manuscript.

Manuscript received 14 April, 2020; final version accepted 17 May, 2021; available online 5 August, 2021.