*Supplement to*

# Statistical Inference on Panel Data Models: A Kernel Ridge Regression Method

This supplement document contains proofs and other relevant results that were not included in the main text and appendix. In Section S.1, we prove Lemmas A.1 and A.2, Propositions A.1 and A.2. In Section S.2, we prove Lemmas A.3, A.4, A.5 A.6, A.7 and A.8. We also provide additional Lemmas S.1, S.2 and S.3 as well as their proofs. Lemmas S.1 and S.2 give mild conditions to guarantee the validity of Assumption A4; Lemma S.3 is useful for proving Lemma A.4.

## S.1. Additional Proofs or Other Relevant Results for Heterogeneous Model

*Proof of Lemma A.1.* For any $\theta, \theta_k = (\beta_k, g_k) \in \Theta_i$ for $k = 1, 2$, it holds from (2.1) that

$$
\begin{aligned}
& \langle DS^{\star}_{i,M,\eta_i}(\theta)\theta_1, \theta_2\rangle_i \\
= & \langle E\{DS_{i,M,\eta_i}(\theta)\theta_1\}, \theta_2\rangle_i \\
= & \frac{1}{T}\sum_{t=1}^{T} E\left(\langle R_iU_{it}, \theta_1\rangle_i \langle R_iU_{it}, \theta_2\rangle_i\right) + \langle P_i\theta_1, \theta_2\rangle_i \\
= & \frac{1}{T}\sum_{t=1}^{T} E\left((g_1(X_{it}) + Z_t'\beta_1)(g_2(X_{it}) + Z_t'\beta_2)\right) + \eta_i\langle g_1, g_2\rangle_{\mathcal{H}_i} \\
= & E\left((g_1(X_i) + Z'\beta_1)(g_2(X_i) + Z'\beta_2)\right) + \eta_i\langle g_1, g_2\rangle_{\mathcal{H}_i} \\
= & \langle \theta_1, \theta_2\rangle_i,
\end{aligned}
$$

which implies that $DS^{\star}_{i,M,\eta_i}(\theta) = id$, the identity operator on $\Theta_i$. □

*Proof of Lemma A.2.* It follows by Proposition 2.1 that

$$
\|R_iU_{it}\|_i^2 = K^{(i)}(X_{it}, X_{it}) + (Z_t - A_i(X_{it}))'(\Omega_i + \Sigma_i)^{-1}(Z_t - A_i(X_{it})).
$$

By (2.6) and $\langle A_i, g\rangle_{\star,i} = V_i(G_i, g)$ (see Section 3.2),

$$
\begin{aligned}
A_i(x) &= \langle A_i, K_x^{(i)}\rangle_{\star,i} = V_i(G_i, K_x^{(i)}) \\
&= \sum_{\nu=1}^{\infty} \frac{\varphi_\nu^{(i)}(x)}{1 + \eta_i\rho_\nu^{(i)}} V_i(G_i, \varphi_\nu^{(i)}).
\end{aligned}
$$

It follows by Assumption A3 that $C_{\varphi,i} \equiv \sup_{\nu\geq 1}\sup_{x\in\mathcal{X}_i}|\varphi_\nu^{(i)}(x)| < \infty$. Then we have

$$\begin{aligned}
K^{(i)}(X_{it}, X_{it}) &= \sum_{\nu\geq 1}\frac{|\varphi_\nu^{(i)}(X_{it})|^2}{1+\eta_i\rho_\nu^{(i)}} \leq C_{\varphi,i}^2 h_i^{-1}, \\
A_i(X_{it})'(\Omega_i+\Sigma_i)^{-1}A_i(X_{it}) &\leq c_1^{-1}A_i(X_{it})'A_i(X_{it}) \\
&\leq c_1^{-1}\sum_{\nu\geq 1}\frac{|\varphi_\nu^{(i)}(X_i)|^2}{(1+\eta_i\rho_\nu^{(i)})^2}\sum_{\nu\geq 1}V_i(G_i', \varphi_\nu^{(i)})V_i(G_i, \varphi_\nu^{(i)}) \\
&\leq c_1^{-1}C_{\varphi,i}^2 C_{G_i}^2 h_i^{-1}, \\
Z_t'(\Omega_i+\Sigma_i)^{-1}Z_t &\leq c_1^{-1}Z_t'Z_t,
\end{aligned}$$

where $C_{G_i}^2 = \sum_{\nu\geq 1}V_i(G_i', \varphi_\nu^{(i)})V_i(G_i, \varphi_\nu^{(i)})$. By Assumption A1, $C_{G_i}^2$ is a finite positive constant. Then (A.1) holds for $C_i^2 = \max\{C_{\varphi,i}^2, 2c_1^{-1}C_{\varphi,i}^2 C_{G_i}^2, 2c_1^{-1}\}$.

To show (A.2), first notice that, for any $\theta = (\beta, g) \in \Theta_i$,

$$\|\theta\|_{i,\sup} = \sup_{x\in\mathcal{X}_i, \|z\|_2=1}|g(x)+z'\beta|.$$

The "$\geq$" is obvious. To show "$\leq$", note that for any $x \in \mathcal{X}_i$, choose $z_x = \text{sign}(g(x))\beta/\|\beta\|_2$. Then

$$|g(x)+z_x'\beta| = |g(x)| + \|\beta\|_2.$$

Therefore,

$$\sup_{x\in\mathcal{X}_i, \|z\|_2=1}|g(x)+z'\beta| \geq \sup_{x\in\mathcal{X}_i}|g(x)+z_x'\beta| = \sup_{x\in\mathcal{X}_i}|g(x)| + \|\beta\|_2 = \|\theta\|_{i,\sup}.$$

Following Proposition 2.1 and the proof of (A.1), for $u = (x, z)$ with $x \in \mathcal{X}_i$ and $\|z\|_2 = 1$,

$$|g(x)+z'\beta| = |\langle R_i u, \theta\rangle_i| \leq \|R_i u\|_i\|\theta\|_i \leq C_i(1+h_i^{-1/2})\|\theta\|_i.$$

This proves (A.2). □

*Proof of Proposition A.1.* Since $f_{1t}$, $f_{2t}$, $v_{it}$ and $\epsilon_{it}$ all have finite $\alpha$th moments, it follows by (2.2) that $X_{it}$ and $Z_t$ both have finite $\alpha$th moments, i.e., $E(\|X_i\|_2^\alpha) < \infty$ and $E(\|Z\|_2^\alpha) < \infty$. Define

$$C_T(\xi) = \inf\{x | TP(\|Z\|_2 > x) \leq \xi\}, \quad \xi > 0.$$

By Markov inequality,

$$P\left(\|Z\|_2 > [TE(\|Z\|_2^\alpha)/\xi]^{1/\alpha}\right) \leq \frac{\xi}{T},$$

therefore,

$$C_T(\xi) \leq \left(\frac{TE(\|Z\|_2^\alpha)}{\xi}\right)^{1/\alpha}.$$

Thanks to the $\phi$-mixing condition (see Assumption A2), it follows by O'Brien (1974, Theorem 1) that for any $\xi > 0$,

$$\liminf_{T\to\infty} P\left(\max_{1\leq t\leq T}\|Z_t\|_2 \leq C_T(\xi)\right) = \exp(-b\xi),$$

where $b > 0$ is a constant. For arbitrary $\varepsilon > 0$, choose $\xi > 0$ such that $1 - \exp(-b\xi) < \varepsilon/2$. Then, as $T$ approaches infinity,

$$P\left(\max_{1\le t\le T} \|Z_t\|_2 \le C_T(\xi)\right) \ge \exp(-b\xi) - \varepsilon/2,$$

leading to that

$$P\left(\max_{1\le t\le T} \|Z_t\|_2 > C_T(\xi)\right) \le 1 - \exp(-b\xi) + \varepsilon/2 \le \varepsilon.$$

This proves that

$$\max_{1\le t\le T} \|Z_t\|_2 = O_P(C_T(\xi)) = O_P(T^{1/\alpha}).$$

□

*Proof of Proposition A.2.* For notation simplicity, denote

$$\begin{aligned} \mathcal{F}_j &= \mathcal{F}_1^j, \quad j \in [T], \\ \mathcal{F}_0 &= \text{trivial } \sigma\text{-algebra consisting only of the empty set and full sample space.} \end{aligned}$$

For any $\theta_1, \theta_2 \in \Theta_i$, define $l_{it} = (\psi_{i,M,t}(U_{it};\theta_1) - \psi_{i,M,t}(U_{it};\theta_2))R_iU_{it}$. First of all, we will prove the following concentration inequality: for any $r > 0$,

$$P\left(\left\|\sum_{t=1}^T [l_{it} - E(l_{it})]\right\|_i \ge r\right) \le 2\exp\left(-\frac{r^2}{32TC_\phi^2\|\theta_1-\theta_2\|_{i,\sup}^2}\right), \tag{S.1}$$

where $C_\phi \equiv \sum_{t=0}^\infty \phi(t)$. It follows by Assumption A2 that $C_\phi$ is finite. Clearly, (S.1) holds for $\theta_1 = \theta_2$ since both sides equal to zero. In what follows, we assume $\theta_1 \ne \theta_2$.

Define $\mathbb{M}_{iT} = \sum_{t=1}^T l_{it}$, and $f_{iTj} = E(\mathbb{M}_{iT}|\mathcal{F}_j) - E(\mathbb{M}_{iT}|\mathcal{F}_{j-1})$, $j \in [T]$. It is easy to see that

$$\begin{aligned} \mathbb{M}_{iT} - E(\mathbb{M}_{iT}) &= \sum_{j=1}^T f_{iTj}, \\ f_{iTj} &= \sum_{t=j}^T \left(E(l_{it}|\mathcal{F}_j) - E(l_{it}|\mathcal{F}_{j-1})\right). \end{aligned} \tag{S.2}$$

Clearly, $f_{iTj}$ is $\mathcal{F}_j$-measurable. For $k \in [T]$, define $\mathbb{N}_{iTk} = \sum_{j=1}^k f_{iTj}$ and $\mathbb{N}_{iT0} \equiv 0$. Then $\mathbb{N}_{iTk} = \mathbb{N}_{iTk-1} + f_{iTk}$. For $\lambda > 0$, let $u_{k-1}(x) = \lambda\|\mathbb{N}_{iTk-1} + xf_{iTk}\|_i$, $x \in [0,1]$. Define

$$\varphi_{k-1}(x) = E\left(\cosh\left(u_{k-1}(x)\right)|\mathcal{F}_{k-1}\right), \ x \in [0,1].$$

It is easy to see that

$$\begin{aligned} \varphi_{k-1}(1) &= E\left(\cosh\left(\lambda\|\mathbb{N}_{iTk}\|_i\right)|\mathcal{F}_{k-1}\right) \\ \varphi_{k-1}(0) &= E\left(\cosh\left(\lambda\|\mathbb{N}_{iTk-1}\|_i\right)|\mathcal{F}_{k-1}\right). \end{aligned}$$

By the proof of Pinelis (1994, Theorem 3.2) and direct calculations, it can be shown that

$$\begin{aligned}\varphi'_{k-1}(x) &= E\left(\sinh(u_{k-1}(x))u'_{k-1}(x)|\mathcal{F}_{k-1}\right),\\ \varphi''_{k-1}(x) &= E\left(\cosh(u_{k-1})(u'_{k-1}(x))^2+\sinh(u_{k-1}(x))u''_{k-1}(x)|\mathcal{F}_{k-1}\right)\\ &\le E\left(\cosh(u_{k-1})(u'_{k-1}(x))^2+\cosh(u_{k-1}(x))u_{k-1}(x)u''_{k-1}(x)|\mathcal{F}_{k-1}\right)\\ &= \frac{1}{2}E\left(\cosh(u_{k-1}(x))(u_{k-1}(x)^2)''|\mathcal{F}_{k-1}\right)\\ &= \lambda^2 E\left(\cosh(u_{k-1}(x))\|f_{iTk}\|_i^2|\mathcal{F}_{k-1}\right).\end{aligned} \tag{S.3}$$

Next we will show that $\|f_{iTk}\|_i^2$ is almost surely bounded. We will first examine the terms $E(l_{it}|\mathcal{F}_k)-E(l_{it})$ for $t\ge k$. Arbitrarily choose $A\in\mathcal{F}_k$ and $\theta\in\Theta_i$ with $\|\theta\|_i=1$. Define $X=\langle l_{it},\theta\rangle_i$. Write $X=X^+-X^-$, where $X^+$ and $X^-$ represent the positive and negative parts of $X$, respectively. Clearly, $|X|\le\|l_{it}\|_i\|\theta\|_i=\|l_{it}\|_i$ implying that both $X^+$ and $X^-$ belong to $[0,\|l_{it}\|_i]$. Note that the $X^+$ is $\mathcal{F}_t^\infty$-measurable. Therefore,

$$\begin{aligned}|E(X^+|A)-E(X^+)| &= \left|\int_0^{\|l_{it}\|_i}[P(X^+>v|A)-P(X^+>v)]dv\right|\\ &\le \int_0^{\|l_{it}\|_i}|P(X^+>v|A)-P(X^+>v)|dv\le\|l_{it}\|_i\phi(t-k).\end{aligned}$$

Similarly, one can show that $|E(X^-|A)-E(X^-)|\le\|l_{it}\|_i\phi(t-k)$. Therefore,

$$|E(X|A)-E(X)|\le 2\|l_{it}\|_i\phi(t-k).$$

By arbitrariness of $A\in\mathcal{F}_k$ and by taking supremum over $\theta\in\Theta_i$ with $\|\theta\|_i=1$, one gets that

$$\|E(l_{it}|\mathcal{F}_k)-E(l_{it})\|_i\le 2\|l_{it}\|_i\phi(t-k),\ t\ge k. \tag{S.4}$$

Similar arguments lead to

$$\|E(l_{it}|\mathcal{F}_{k-1})-E(l_{it})\|_i\le 2\|l_{it}\|_i\phi(t-k+1),\ t\ge k.$$

Therefore, for $t\ge k$,

$$\|E(l_{it}|\mathcal{F}_k)-E(l_{it}|\mathcal{F}_{k-1})\|_i\le 2\|l_{it}\|_i(\phi(t-k)+\phi(t-k+1)).$$

Using (S.2) and the assumption $\|l_{it}\|_i\le\|\theta_1-\theta_2\|_{i,\sup}$, it can be shown that

$$\begin{aligned}\|f_{iTk}\|_i &\le \sum_{t=k}^T\|E(l_{it}|\mathcal{F}_k)-E(l_{it}|\mathcal{F}_{k-1})\|_i\\ &\le \sum_{t=k}^T 2\|l_{it}\|_i(\phi(t-k)+\phi(t-k+1))\\ &\le 2\|\theta_1-\theta_2\|_{i,\sup}\sum_{t=k}^T(\phi(t-k)+\phi(t-k+1))\\ &\le 4C_\phi\|\theta_1-\theta_2\|_{i,\sup}.\end{aligned}$$

Therefore, it follows by (S.3) that

$$\varphi''_{k-1}(x) \le 16\lambda^2 C_\phi^2 \|\theta_1 - \theta_2\|_{i,\sup}^2 \varphi_{k-1}(x).$$

Meanwhile, note that $\mathbb{N}_{iTk-1}$ is $\mathcal{F}_{k-1}$-measurable, so we have

$$\varphi'_{k-1}(0) = \lambda \frac{\sinh(\lambda \|\mathbb{N}_{iTk-1}\|_i)}{\|\mathbb{N}_{iTk-1}\|_i} E\left(\langle \mathbb{N}_{iTk-1}, f_{iTk}\rangle_i | \mathcal{F}_{k-1}\right) = 0, \ k \ge 2,$$

where the last equality follows from $E(f_{iTk}|\mathcal{F}_{k-1}) = 0$. Directly using (S.3) one also has that $\varphi'_0(0) = 0$. So $\varphi'_{k-1}(0) = 0$ for all $k \in [T]$. By Dudley et al. (1992, pp. 133, Lemma 3) we have for $k \in [T]$,

$$\varphi_{k-1}(x) \le \varphi_{k-1}(0) \exp\left(8\lambda^2 C_\phi^2 \|\theta_1 - \theta_2\|_{i,\sup}^2 x^2\right), \quad x \in [0, 1].$$

In particular,

$$\varphi_{k-1}(1) \le \varphi_{k-1}(0) \exp\left(8\lambda^2 C_\phi^2 \|\theta_1 - \theta_2\|_{i,\sup}^2\right).$$

Taking expectations on both sides leading to that

$$E\left(\cosh\left(\lambda \|\mathbb{N}_{iTk}\|_i\right)\right) \le \exp\left(8\lambda^2 C_\phi^2 \|\theta_1 - \theta_2\|_{i,\sup}^2\right) E\left(\cosh\left(\lambda \|\mathbb{N}_{iTk-1}\|_i\right)\right). \tag{S.5}$$

By repeatedly using (S.5) and the convention $\mathbb{N}_{iT0} = 0$, and by (S.2), we have

$$\begin{aligned} E\left(\cosh\left(\lambda \left\| \sum_{t=1}^T [l_{it} - E(l_{it})] \right\|_i\right)\right) &= E\left(\cosh\left(\lambda \|\mathbb{N}_{iTT}\|_i\right)\right) \\ &\le \exp\left(8T\lambda^2 C_\phi^2 \|\theta_1 - \theta_2\|_{i,\sup}^2\right). \end{aligned}$$

Therefore,

$$\begin{aligned} P\left(\left\| \sum_{t=1}^T [l_{it} - E(l_{it})] \right\|_i \ge r\right) &= P\left(\cosh\left(\lambda \left\| \sum_{t=1}^T [l_{it} - E(l_{it})] \right\|_i\right) \ge \cosh(\lambda r)\right) \\ &\le \frac{1}{\cosh(\lambda r)} E\left(\cosh\left(\lambda \left\| \sum_{t=1}^T [l_{it} - E(l_{it})] \right\|_i\right)\right) \\ &\le e \exp\left(-\lambda r + 8T\lambda^2 C_\phi^2 \|\theta_1 - \theta_2\|_{i,\sup}^2\right). \end{aligned}$$

Then (S.1) follows by choosing

$$\lambda = \frac{r}{16 T C_\phi^2 \|\theta_1 - \theta_2\|_{i,\sup}^2}.$$

The rest of the proof follows by chaining argument. Let $\psi_2(x) = \exp(x^2) - 1$. It follows by (S.1) and Kosorok (2008, Theorem 8.1) that for any $\theta_1, \theta_2 \in \Theta_i$,

$$\left\| \|\mathbb{Z}_{iM}(\theta_1) - \mathbb{Z}_{iM}(\theta_2)\|_i \right\|_{\psi_2} \le \sqrt{96} C_\phi \|\theta_1 - \theta_2\|_{i,\sup}.$$

It follows by Kosorok (2008, Theorem 8.4) that there exists a universal constant $C > 0$, which only depends on $C_\psi$, such that for any $\delta > 0$,

$$\begin{aligned}
&\left\| \sup_{\substack{\theta_1,\theta_2\in\mathcal{G}_i(p_i)\\ \|\theta_1-\theta_2\|_{i,\sup}\le\delta}} \|\mathbb{Z}_{iM}(\theta_1)-\mathbb{Z}_{iM}(\theta_2)\|_i \right\|_{\psi_2}\\
\le\ & C\left(\int_0^\delta \psi_2^{-1}\left(D_i(\varepsilon,\mathcal{G}_i(p_i),\|\cdot\|_{i,\sup})\right)d\varepsilon + \delta\psi_2^{-1}\left(D_i(\delta,\mathcal{G}_i(p_i),\|\cdot\|_{i,\sup})^2\right)\right) = CJ_i(p_i,\delta).
\end{aligned}$$

Therefore,

$$\left\| \sup_{\substack{\theta\in\mathcal{G}_i(p_i)\\ \|\theta\|_{i,\sup}\le\delta}} \|\mathbb{Z}_{iM}(\theta)\|_i \right\|_{\psi_2} \le CJ_i(p_i,\delta).$$

It follows again from Kosorok (2008, Lemma 8.1) that for all $\delta > 0, s > 0$,

$$P\left(\sup_{\theta\in\mathcal{G}_i(p_i),\|\theta\|_{i,\sup}\le\delta} \|\mathbb{Z}_{iM}(\theta)\|_i > s\right) \le 2\exp\left(-\frac{s^2}{C^2J_i(p_i,\delta)^2}\right). \tag{S.6}$$

It is easy to see that for any $\theta\in\mathcal{G}_i(p_i)$, $\|\theta\|_{i,\sup}\le 1$. Let $\sqrt{T}J_i(p_i,1)=\varepsilon^{-1}$, and $Q_\varepsilon = -\log\varepsilon - 1$. Let $\tau = 3C\sqrt{\log N + \log\log(TJ_i(p_i,1))}$. Then it can be checked that

$$N(Q_\varepsilon+2)\exp\left(-\frac{\tau^2}{C^2\exp(2)}\right)\to 0, \quad N\to\infty.$$

Since $J_i(p_i,\delta)$ is strictly increasing in $\delta$, the function $J_i(\delta)\equiv J_i(p_i,\delta)$ has inverse denoted by $J_i^{-1}$. Then we have

$$\begin{aligned}
& P\left(\max_{i\in[N]}\sup_{\theta\in\mathcal{G}_i(p_i)} \frac{\sqrt{T}\|\mathbb{Z}_{iM}(\theta)\|_i}{\sqrt{T}J_i(\|\theta\|_{i,\sup})+1}\ge\tau\right)\\
\le\ & \sum_{i=1}^N\left(P\left(\sup_{\|\theta\|_{i,\sup}\le J_i^{-1}(T^{-1/2})} \frac{\sqrt{T}\|\mathbb{Z}_{iM}(\theta)\|_i}{\sqrt{T}J_i(\|\theta\|_{i,\sup})+1}\ge\tau\right)\right.\\
&\left. + \sum_{l=0}^{Q_\varepsilon} P\left(\sup_{J_i^{-1}(T^{-1/2}\exp(l))\le\|\theta\|_{i,\sup}\le J_i^{-1}(T^{-1/2}\exp(l+1))} \frac{\sqrt{T}\|\mathbb{Z}_{iM}(\theta)\|_i}{\sqrt{T}J_i(\|\theta\|_{i,\sup})+1}\ge\tau\right)\right)\\
\le\ & \sum_{i=1}^N\left(P\left(\sup_{\|\theta\|_{i,\sup}\le J_i^{-1}(T^{-1/2})} \|\mathbb{Z}_{iM}(\theta)\|_i\ge T^{-1/2}\tau\right)\right.\\
&\left. + \sum_{l=0}^{Q_\varepsilon} P\left(\sup_{\|\theta\|_{i,\sup}\le J_i^{-1}(T^{-1/2}\exp(l+1))} \|\mathbb{Z}_{iM}(\theta)\|_i\ge T^{-1/2}\tau\exp(l)\right)\right)\\
\le\ & \sum_{i=1}^N\left(2\exp\left(-\tau^2/C^2\right) + \sum_{l=0}^{Q_\varepsilon} 2\exp\left(-\tau^2/(C^2\exp(2))\right)\right)\\
=\ & 2N(Q_\varepsilon+2)\exp\left(-\frac{\tau^2}{C^2\exp(2)}\right)\to 0, \quad \text{as } N\to\infty.
\end{aligned} \tag{S.7}$$

This proves the desirable conclusion with $C_0 = 3C$. □

## S.2. Additional Proofs or Other Relevant Results for Homogeneous Model

The following lemma gives mild conditions that guarantee Assumption A4. Before stating the lemma, we borrow the concept of complete continuity from Weinberger (1974, page 50). A bilinear functional $A(\cdot,\cdot)$ on $\mathcal{H}\times\mathcal{H}$ is said to be completely continuous w.r.t another bilinear functional $B(\cdot,\cdot)$ if for any $\epsilon>0$, there exists finite number of functionals $l_1, l_2, ..., l_k$ on $\mathcal{H}$ such that $l_i(g)=0, i=1,2,..,k$, implies $A(g,g)\le \epsilon B(g,g)$.

Let $U$ be an open subset of $\mathcal{X}$ and $U^{NT}\equiv \underbrace{U\times U\times\cdots\times U}_{NT \text{ items}}$. Let $C(\mathcal{X})$ be the set of all continuous functions on $\mathcal{X}$ and $\mathcal{H}\subseteq C(\mathcal{X})$. Let $\mathbf{x}$ denote the $NT$-vector $(x_{11},\dots,x_{1T},\dots,x_{N1},\dots,x_{NT})$.

**Lemma S.1.** *Suppose $1\notin\mathcal{H}$, and $p(\mathbf{x}|\mathcal{F}_1^T)>0$ for $\mathbf{x}\in U^{NT}$, where $p(\mathbf{x}|\mathcal{F}_1^T)$ is the joint conditional density of $X_{11}, X_{12}, ..., X_{NT}$ given $\mathcal{F}_1^T$. If $V(f,g)=0$ for all $f\in\mathcal{H}$, then $g=0$.*

*Proof of Lemma S.1.* For simplicity, we assume that $f_{1t}, X_{it}$ are both univariate. By assumption, $0=V(g,g)=\sum_{i=1}^N E\left\{(\tau_i g)'P(\tau_i g)\middle|\mathcal{F}_1^T\right\}/(NT)$. Hence it follows that

$$0=\int (g(x_{i1}),g(x_{i2}),...,g(x_{iT}))P(g(x_{i1}),g(x_{i2}),...,g(x_{iT}))'p(\mathbf{x}|\mathcal{F}_1^T)d\mathbf{x}, \text{ for all } i\in[N]. \tag{S.1}$$

Since the integrand in (S.1) is continuous and nonnegative, it holds that, for all $i\in[N]$ and $\mathbf{x}$ with $p(\mathbf{x}|\mathcal{F}_1^T)>0$,

$$(g(x_{i1}),g(x_{i2}),...,g(x_{iT}))P(g(x_{i1}),g(x_{i2}),...,g(x_{iT}))'=0. \tag{S.2}$$

By definition, $P$ is a projection matrix whose image is the orthogonal space of the linear space spanned by $F_1$ and $\bar{X}$. Therefore, it yields that

$$(g(x_{i1}),g(x_{i2}),...,g(x_{iT}))=\alpha_i(f_{11},f_{12},...,f_{1T})+\beta_i(\bar{x}_1,\bar{x}_2,...,\bar{x}_T), \tag{S.3}$$

for some $\alpha_i,\beta_i\in\mathbb{R}$. Consider $\mathbf{x}=(x_{11},x_{12},...,x_{N,T-1},x_{NT})$ and $\widetilde{\mathbf{x}}=(x_{11},x_{12},...,x_{N,T-1},\widetilde{x}_{NT})\in U^{NT}$ with $x_{NT}\neq\widetilde{x}_{NT}$ and $p(\mathbf{x}|\mathcal{F}_1^T)>0, p(\widetilde{\mathbf{x}}|\mathcal{F}_1^T)>0$, i.e., the two points differ only on the last element. Applying (S.2) to point $\widetilde{\mathbf{x}}$, we have

$$(g(x_{i1}),g(x_{i2}),...,g(\widetilde{x}_{iT}))=\widetilde{\alpha}_i(f_{11},f_{12},...,f_{1T})+\widetilde{\beta}_i(\bar{x}_1,\bar{x}_2,...,\widetilde{\bar{x}}_T), \tag{S.4}$$

for some $\widetilde{\alpha}_i,\widetilde{\beta}_i\in\mathbb{R}$. Comparing (S.3) and (S.4), and by the fact $T>q_1+d=2$, it holds that $\alpha_i=\widetilde{\alpha}_i, \beta_i=\widetilde{\beta}_i=0$. Hence $(g(x_{i1}),g(x_{i2}),...,g(x_{iT}))\in span((f_{11},f_{12},...,f_{1T}))$ for all $p(\mathbf{x}|\mathcal{F}_1^T)>0, i\in[N]$, and it happens if and only if $g=0$. ☐

**Lemma S.2.** *Suppose $\mathcal{X}$ is compact. Furthermore if $V(f,g)=0$ for all $f\in\mathcal{H}$ implies $g=0$, then Assumption A4 is valid.*

**Remark S.2.1.** *The compactness of $\mathcal{X}$ can be relaxed by Mercer's theorem; see Sun (2005).*

*Proof of Lemma S.2.* Define bilinear functionals $W(g,\widetilde{g}) = \sum_{i=1}^N E\left\{(\tau_i g)'(\tau_i \widetilde{g})\middle|\mathcal{F}_1^T\right\}/(NT)$, and $J(g,\widetilde{g}) = \langle g,\widetilde{g}\rangle_{\mathcal{H}}$. Clearly, $V(g,g) \le W(g,g)$. Let $\mu$ be a measure such that

$$\int g d\mu = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T E(g(X_{it})|\mathcal{F}_1^T).$$

Hence, $\int g^2 d\mu = W(g,g)$. By Mercer's theorem, the kernel $\bar{K}$ of $\mathcal{H}$ follows the expansion:

$$\bar{K}(x,y) = \sum_{i=1}^\infty \lambda_i e_i(x)e_i(y).$$

where $\lambda_i$ is a non-increasing positive sequence converging to zero and $\{e_i\}_{i=1}^\infty$ forms an orthonormal basis of $L_2(\mu)$, so that $W(e_i,e_j) = \delta_{ij}$. Moreover, $\{\sqrt{\lambda_i}e_i\}_{i=1}^\infty$ is also an orthonormal basis of $\mathcal{H}$, which is proved in Cucker and Smale (2001). As a consequence, any $g\in\mathcal{H}$ simultaneously admits the following expansions:

$$g = \sum_{i=1}^\infty W(g,e_i)e_i, \quad g = \sum_{i=1}^\infty J(g,\sqrt{\lambda_i}e_i)\sqrt{\lambda_i}e_i$$

with $\sum_{i=1}^\infty W^2(g,e_i) < \infty$ and $\sum_{i=1}^\infty J^2(g,\sqrt{\lambda_i}e_i) < \infty$. This implies $W(g,e_i) = \lambda_i J(g,e_i)$. For any $\epsilon > 0$, choose integer $k$ large enough so that $\lambda_i < \epsilon$ for $i > k$. Define functionals $l_i(g) = W(g,e_i), i = 1,2,...,k$. By direct direct examinations, if $l_i(g) = 0$ for $i = 1,2,...,k$, then

$$W(g,g) = \sum_{i=k+1}^\infty W^2(g,e_i) = \sum_{i=k+1}^\infty \lambda_i^2 J^2(g,e_i) \le \epsilon\sum_{i=k+1}^\infty \lambda_i J^2(g,e_i) = \epsilon J(g,g).$$

Since $V(g,g) \le W(g,g) \le \epsilon J(g,g)$, $V$ is completely continuous w.r.t $J$. By Weinberger (1974, Theorem 3.1, page 52), there are positive eigenvalues $\{\alpha_i\}_{i=1}^\infty$ converging to zero and eigenfunctions $\{\widetilde{\varphi}_i\}_{i=1}^\infty \in \mathcal{H}$ such that $V(\widetilde{\varphi}_i,\widetilde{\varphi}_j) = \alpha_i\delta_{ij}$, $J(\widetilde{\varphi}_i,\widetilde{\varphi}_j) = \delta_{ij}$ and

$$g = \sum_{i=1}^\infty J(g,\widetilde{\varphi}_i)\widetilde{\varphi}_i, \text{ for all } g\in\mathcal{H}.$$

The above implies $V(g,\widetilde{\varphi}_i) = \alpha_i J(g,\widetilde{\varphi}_i)$. Take $\varphi_i = \widetilde{\varphi}_i/\sqrt{\alpha_i}$ and $\rho_i = 1/\alpha_i$, then $\{\varphi_i\}_{i=1}^\infty$ and $\{\rho_i\}_{i=1}^\infty$ will satisfy Assumption A4. □

*Proof of Lemma A.3.* Throughout we let $\|A\|_F = \sqrt{\mathrm{Tr}(AA')}$ be the Frobenius norm. Clearly,

$$\widetilde{\Sigma} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \bar{v}_1 & \bar{v}_2 & \cdots & \bar{v}_T \end{pmatrix}, \ \Sigma_\star\widetilde{\Sigma}' = \left(0_{(q_1+d)\times q_1}, \sum_{t=1}^T Z_t^\star \bar{v}_t'\right).$$

By direct examinations we have

$$\Sigma\Sigma' - \Sigma_\star\Sigma_\star' \quad = \quad \Sigma_\star\widetilde{\Sigma}' + \widetilde{\Sigma}'\Sigma_\star + \widetilde{\Sigma}\widetilde{\Sigma}' \equiv R. \tag{S.5}$$

By independence of $v_{it}$ and $Z_t^\star$, it can be shown that

$$
\begin{aligned}
&E\left(\|\sum_{t=1}^T Z_t^\star \bar{v}_t'\|_F^2\right) = \sum_{t,l=1}^T \mathrm{Tr}\left(E\left(\bar{v}_t'\bar{v}_l\right) E\left((Z_t^\star)' Z_l^\star\right)\right) = O(T/N),\\
&E\left(\|\widetilde{\Sigma}\widetilde{\Sigma}'\|_F^2\right) \le E\left(\mathrm{Tr}\left(\widetilde{\Sigma}\widetilde{\Sigma}'\widetilde{\Sigma}\widetilde{\Sigma}'\right)\right) = O(T^2/N^2). \qquad \text{(S.6)}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
E\left(\|\Sigma_\star \widetilde{\Sigma}'\|_F^2\right) &= E\left(\|\sum_{t=1}^T Z_t^\star \bar{v}_t'\|_F^2\right) = O(T/N),\\
E(\|R\|_F^2) &\le 8E(\|\Sigma_\star\widetilde{\Sigma}'\|_F^2) + 2E(\|\widetilde{\Sigma}\widetilde{\Sigma}'\|_F^2) = O(T/N + (T/N)^2). \qquad \text{(S.7)}
\end{aligned}
$$

Since

$$
\begin{aligned}
\|(\Sigma\Sigma')^{-1} - (\Sigma_\star\Sigma_\star')^{-1}\|_{\mathrm{op}} &= \|(\Sigma_\star\Sigma_\star')^{-1} R (\Sigma\Sigma')^{-1}\|_{\mathrm{op}}\\
&\le \|(\Sigma_\star\Sigma_\star')^{-1}\|_{\mathrm{op}} \|R\|_{\mathrm{op}} \|(\Sigma\Sigma')^{-1}\|_{\mathrm{op}},
\end{aligned}
$$

it follows by Assumption A5 and (S.5) and Hölder inequality that

$$
E\left(\|(\Sigma\Sigma')^{-1} - (\Sigma_\star\Sigma_\star')^{-1}\|_{\mathrm{op}}^{1+\omega}\right) = O((T^3N)^{-(1+\omega)/2} + (TN)^{-(1+\omega)}), \qquad \text{(S.8)}
$$

where $\omega = (\zeta-4)/(\zeta+4)$. Note that $E(\widetilde{\Sigma}'\widetilde{\Sigma}) = \sigma_v^2 I_T$ and $E\mathrm{Tr}(\widetilde{\Sigma}'\widetilde{\Sigma}) = O(T/N)$, where $\sigma_v^2 = E(v_{it}'v_{it})$ is a constant. By direct examinations

$$
\begin{aligned}
& P_\star - P\\
=\ & \Sigma'\left((\Sigma\Sigma')^{-1} - (\Sigma_\star\Sigma_\star')^{-1}\right)\Sigma + \Sigma_\star'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma} + \widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\Sigma_\star + \widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma}.
\end{aligned}
$$

It follows by (S.6), (S.7) and (S.8) and Hölder inequality that

$$
\begin{aligned}
& E\left(\|P - P_\star\|_{\mathrm{op}}\right)\\
\le\ & E\left(\|\Sigma\Sigma'\|_{\mathrm{op}}\|(\Sigma\Sigma')^{-1} - (\Sigma_\star\Sigma_\star')^{-1}\|_{\mathrm{op}}\right) + 2E\left(\|\widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\Sigma_\star\|_{\mathrm{op}}\right) + E\left(\|\widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma}\|_{\mathrm{op}}\right)\\
\le\ & E\left(\|(\Sigma\Sigma')^{-1} - (\Sigma_\star\Sigma_\star')^{-1}\|_{\mathrm{op}}^{1+\omega}\right)^{1/(1+\omega)} E\left(\|\Sigma\Sigma'\|_{\mathrm{op}}^{(1+\omega)/\omega}\right)^{\omega/(1+\omega)}\\
& +2E\left(\|(\Sigma_\star\Sigma_\star')^{-1}\|_{\mathrm{op}}\right)^{1/2} E\left(\mathrm{Tr}\left(\widetilde{\Sigma}\widetilde{\Sigma}'\right)\right)^{1/2} + E\left(\|(\Sigma_\star\Sigma_\star')^{-1}\|_{\mathrm{op}}\right) E\left(\mathrm{Tr}\left(\widetilde{\Sigma}\widetilde{\Sigma}'\right)\right)\\
=\ & O((T^3N)^{-1/2} + (TN)^{-1})T + O(T^{-1/2}\sqrt{T/N}) + O(T^{-1}(T/N)) = O(N^{-1/2}).
\end{aligned}
\qquad \text{(S.9)}
$$

This proves (A.20). Next we show (A.21). For any $i \in [N]$,

$$
\begin{aligned}
& E\{\gamma_{2i}'F_2'(P - P_\star)K_{\mathbb{X}_i}|\mathcal{F}_1^T\}\\
=\ & E\{\gamma_{2i}'F_2'\Sigma'[(\Sigma\Sigma')^{-1} - (\Sigma_\star\Sigma_\star')^{-1}]\Sigma K_{\mathcal{X}_i}|\mathcal{F}_1^T\} + E\{\gamma_{2i}'F_2'\Sigma_\star'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma}K_{\mathcal{X}_i}|\mathcal{F}_1^T\}\\
& +E\{\gamma_{2i}'F_2'\widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\Sigma_\star K_{\mathbb{X}_i}|\mathcal{F}_1^T\} + E\{\gamma_{2i}'F_2'\widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma}K_{\mathbb{X}_i}|\mathcal{F}_1^T\}.
\end{aligned}
$$

By direct calculations it can be examined that

$$
\begin{aligned}
& E\left(\|\Sigma'[(\Sigma\Sigma')^{-1}-(\Sigma_\star\Sigma_\star')^{-1}]\Sigma F_2\|_{\rm op}\right)\\
\leq\ & E\left(\|(\Sigma\Sigma')^{-1}-(\Sigma_\star\Sigma_\star')^{-1}\|_{\rm op}\times\|\Sigma\Sigma'\|_{\rm op}\times\|F_2\|_{\rm op}\right)\\
\leq\ & E\left(\|(\Sigma\Sigma')^{-1}-(\Sigma_\star\Sigma_\star')^{-1}\|_{\rm op}^{1+\omega}\right)^{1/(1+\omega)}\\
& \times E\left(\|\Sigma\Sigma'\|_{\rm op}^{2(1+\omega)/\omega}\right)^{\omega/(2(1+\omega))}E\left(\|F_2\|_{\rm op}^{2(1+\omega)/\omega}\right)^{\omega/(2(1+\omega))}\\
=\ & O((T^3N)^{-1/2}+(TN)^{-1})T^{3/2}=O(N^{-1/2}+T^{1/2}/N),
\end{aligned}
$$

and

$$
\begin{aligned}
& E\left(\|\Sigma_\star'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma}F_2\|_{\rm op}\right)\\
\leq\ & E\left(\|(\Sigma_\star\Sigma_\star')^{-1}\|_{\rm op}^{1/2}{\rm Tr}(F_2'\widetilde{\Sigma}'\widetilde{\Sigma}F_2)^{1/2}\right)\\
\leq\ & E\left(\|(\Sigma_\star\Sigma_\star')^{-1}\|_{\rm op}\right)^{1/2}E\left({\rm Tr}(F_2'\widetilde{\Sigma}'\widetilde{\Sigma}F_2)\right)^{1/2}\\
=\ & E\left(\|(\Sigma_\star\Sigma_\star')^{-1}\|_{\rm op}\right)^{1/2}E\left({\rm Tr}(F_2'F_2)\right)^{1/2}O(N^{-1/2})\\
=\ & E\left(\|(\Sigma_\star\Sigma_\star')^{-1}\|_{\rm op}\right)^{1/2}E\left(\|F_2'F_2\|_{\rm op}\right)^{1/2}O(N^{-1/2})=O(N^{-1/2}).
\end{aligned}
$$

For any $g\in\mathcal{H}$ with $\|g\|=1$ (implying $|g(x)|\leq c_\varphi h^{-1/2}$ for any $x$), we have

$$
\|E\{F_2'\Sigma_\star'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma}\tau_i g|\mathcal{F}_1^T\}\|_2\leq\|F_2'\Sigma_\star'(\Sigma_\star\Sigma_\star')^{-1}\|_{\rm op}\|E\{\widetilde{\Sigma}\tau_i g|\mathcal{F}_1^T\}\|_2=O_P(\|E\{\widetilde{\Sigma}\tau_i g|\mathcal{F}_1^T\}\|_2).
$$

On the other hand, by direct examinations we have

$$
\|\widetilde{\Sigma}\tau_i g\|_2^2 = \sum_{t,l=1}^T\bar{v}_t'\bar{v}_l g(x_{it})g(x_{il}).
$$

Meanwhile, for any $t\neq l$, $\bar{v}_t'g(x_{it})$ and $\bar{v}_l g(x_{il})$ are independent conditional on $\mathcal{F}_1^T$, and

$$
E\{\bar{v}_l g(x_{il})|\mathcal{F}_1^T\}=\frac{1}{N}E\{v_{il}g(x_{il})|\mathcal{F}_1^T\}+\frac{1}{N}\sum_{k\neq i}E\{v_{kl}g(x_{il})|\mathcal{F}_1^T\}=\frac{1}{N}E\{v_{il}g(x_{il})|\mathcal{F}_1^T\}.
$$

The last equality holds because $v_{kl}$ and $g(x_{il})$ are conditional independent (on $\mathcal{F}_1^T$) for $k\neq i$ and the former has mean zero. This leads us to that

$$
\begin{aligned}
E\{\|\widetilde{\Sigma}\tau_i g\|_2^2|\mathcal{F}_1^T\} &= \sum_{t=1}^T E\{\bar{v}_t'\bar{v}_t g(x_{it})^2|\mathcal{F}_1^T\}+\sum_{t\neq l}E\{\bar{v}_t'\bar{v}_l g(x_{it})g(x_{il})|\mathcal{F}_1^T\}\\
&= \sum_{t=1}^T E\{\bar{v}_t'\bar{v}_t g(x_{it})^2|\mathcal{F}_1^T\}+\sum_{t\neq l}E\{\bar{v}_t'g(x_{it})|\mathcal{F}_1^T\}E\{\bar{v}_l g(x_{il})|\mathcal{F}_1^T\}\\
&= \sum_{t=1}^T E\{\bar{v}_t'\bar{v}_t g(x_{it})^2|\mathcal{F}_1^T\}+\frac{1}{N^2}\sum_{t\neq l}E\{v_{it}'g(x_{it})|\mathcal{F}_1^T\}E\{v_{il}g(x_{il})|\mathcal{F}_1^T\}\\
&= O_P\left(\frac{T}{Nh}+\frac{T^2}{N^2h}\right).
\end{aligned}
$$

Therefore,

$$\|E\{F_2'\Sigma_\star'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma}\tau_i g|\mathcal{F}_1^T\}\|_2 = O_P\left(\sqrt{\frac{T}{Nh}}+\frac{T}{N\sqrt{h}}\right),$$

where the $O_P$ term is free of $g$.

Similarly, we can show that

$$\begin{aligned}
& E\left(\|\widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma}F_2\|_{\text{op}}\right) \\
\leq\ & E\left(\|\widetilde{\Sigma}\widetilde{\Sigma}'\|_{\text{op}}^2\right)^{1/4} E\left(\|(\Sigma_\star\Sigma_\star')^{-1}\|_{\text{op}}^4\right)^{1/4} E\left(F_2'\widetilde{\Sigma}'\widetilde{\Sigma}F_2\right)^{1/2} \\
=\ & O(\sqrt{T/N})O(1/T)O(\sqrt{T/N}) = O(1/N).
\end{aligned} \tag{S.10}$$

Combining the above, we get that

$$\|E\{\gamma_{2i}'F_2'(P-P_\star)K_{\mathbb{X}_i}|\mathcal{F}_1^T\}\| = O_P\left(\sqrt{\frac{T}{Nh}}+\frac{T}{N\sqrt{h}}\right),$$

where the $O_P$ is free of $i\in[N]$. Proof completed. □

**Lemma S.3.** *Suppose that Assumptions A2, A4 and A5 hold. Let $\psi$ satisfy the conditions in Lemma A.4. Then*

$$\sup_{\|g\|_{\sup}\leq 1}\frac{1}{\sqrt{N}}\|\sum_{i=1}^N\psi(\mathbb{X}_i,g)'(P-P_\star)K_{\mathbb{X}_i}\| = O_P(1),$$

*and*

$$\sup_{\|g\|_{\sup}\leq 1}\frac{1}{\sqrt{N}}\|\sum_{i=1}^N E\left(\psi(\mathbb{X}_i,g)'(P-P_\star)K_{\mathbb{X}_i}|\mathcal{F}_1^T\right)\| = O_P(1).$$

*Proof of Lemma S.3.* For any $g,\widetilde{g}$ satisfying $\|g\|_{\sup}\leq 1$ and $\|\widetilde{g}\|\leq 1$, the former implies that $\|\psi(\mathbb{X}_i,g)\|_2\leq L\sqrt{h/T}$ for each $i\in[N]$, and the latter implies that $\|\widetilde{g}\|_{\sup}\leq c_\varphi h^{-1/2}$, by (A.20) we have

$$\frac{1}{\sqrt{N}}\left|\sum_{i=1}^N\psi(\mathbb{X}_i,g)'(P-P_\star)\tau_i\widetilde{g}\right| \leq \frac{1}{\sqrt{N}}\sum_{i=1}^N\|\psi(\mathbb{X}_i,g)\|_2\|\tau_i\widetilde{g}\|_2\|P-P_\star\|_{\text{op}} = O_P(1),$$

and

$$\frac{1}{\sqrt{N}}\left|\sum_{i=1}^N E\left(\psi(\mathbb{X}_i,g)'(P-P_\star)\tau_i\widetilde{g}|\mathcal{F}_1^T\right)\right| \leq Lc_\varphi\sqrt{N}E\left(\|P-P_\star\|_{\text{op}}|\mathcal{F}_1^T\right) = O_P(1).$$

Proof is completed. □

*Proof of Lemma A.4.* It follows by Lemma S.3 that we only need to consider the process $Z_M^\star(g) = \frac{1}{\sqrt{N}}\sum_{i=1}^N[\psi(\mathbb{X}_i,g)'P_\star K_{\mathbb{X}_i} - E\{\psi(\mathbb{X}_i,g)'P_\star K_{\mathbb{X}_i}|\mathcal{F}_1^T\}]$ for $g\in\mathcal{H}$ where the items in summation are independent conditional on $\mathcal{F}_1^T$. Let $\mathbf{K}_i = [K(X_{it}, X_{il})]_{1\le t,l\le T}$, a $T\times T$ matrix. By Assumption A4 it follows that $\mathbf{K}_i \le c_\varphi^2 h^{-1} T I_T$. For any $g_1, g_2 \in \mathcal{H}$,

$$
\begin{aligned}
& \|(\psi(\mathbb{X}_i, g_1) - \psi(\mathbb{X}_i, g_2))' P_\star K_{\mathbb{X}_i}\|^2 \\
= \;& (\psi(\mathbb{X}_i, g_1) - \psi(\mathbb{X}_i, g_2))' P_\star \mathbf{K}_i P_\star (\psi(\mathbb{X}_i, g_1) - \psi(\mathbb{X}_i, g_2)) \\
\le \;& (Lc_\varphi \|P_\star\|_{\mathrm{op}} \|g_1 - g_2\|_{\sup})^2 = (Lc_\varphi \|g_1 - g_2\|_{\sup})^2.
\end{aligned}
$$

The last equation follows by $\|P_\star\|_{\mathrm{op}} = 1$ since $P_\star$ is idempotent. Notice that $\{X_{it} : i \in [N], t \in [T]\}$ are conditional independent given $\mathcal{F}_1^T$. It follows by Pinelis (1994, Theorem 3.5) that for any $r \ge 0$,

$$
P\left( \|Z_M^\star(g_1) - Z_M^\star(g_2)\| \ge r \middle| \mathcal{F}_1^T \right) \le 2\exp\left( -\frac{r^2}{8L^2 c_\varphi^2 \|g_1 - g_2\|_{\sup}^2} \right).
$$

It follows by Kosorok (2008, Lemma 8.1) that

$$
\left\| \|Z_M^\star(g_1) - Z_M^\star(g_2)\| \right\|_{\mathcal{F}_1^T, \psi_2} \le 5Lc_\varphi \|g_1 - g_2\|_{\sup},
$$

where $\|\cdot\|_{\mathcal{F}_1^T,\psi_2}$ denotes the Orlicz-norm conditional on $\mathcal{F}_1^T$ with respect to $\psi_2(s) = \exp(s^2) - 1$. This in turn leads to, by Kosorok (2008, Theorem 8.4), that for any $\delta > 0$,

$$
\begin{aligned}
& \left\| \sup_{\substack{g_1, g_2 \in \mathcal{G}(p) \\ \|g_1 - g_2\|_{\sup} \le \delta}} \|Z_M^\star(g_1) - Z_M^\star(g_2)\| \right\|_{\mathcal{F}_1^T, \psi_2} \\
\le \;& C\left[ \int_0^\delta \psi_2^{-1}\left( D(\varepsilon, \mathcal{G}(p), \|\cdot\|_{\sup}) \right) d\varepsilon + \delta \psi_2^{-1}\left( D(\delta, \mathcal{G}(p), \|\cdot\|_{\sup})^2 \right) \right] \\
= \;& CJ(p, \delta),
\end{aligned}
$$

where $C > 0$ is a constant depending on $L, c_\varphi$ only. Then we have

$$
\left\| \sup_{\substack{g \in \mathcal{G}(p) \\ \|g\|_{\sup} \le \delta}} \|Z_M^\star(g)\| \right\|_{\mathcal{F}_1^T, \psi_2} \le CJ(p, \delta).
$$

It follows again from Kosorok (2008, Lemma 8.1) that for all $\delta > 0, r > 0$,

$$
P\left( \sup_{\substack{g \in \mathcal{G}(p) \\ \|g\|_{\sup} \le \delta}} \|Z_M^\star(g)\| \ge r \middle| \mathcal{F}_1^T \right) \le 2\exp\left( -\frac{r^2}{C^2 J(p, \delta)^2} \right). \tag{S.11}
$$

Let $Q_N = \log(N^{1/2} J(p, 1)) - 1$. It follows from the proof of (S.7) that

$$
\begin{aligned}
& P\left( \sup_{g \in \mathcal{G}(p)} \frac{\sqrt{N}\|Z_M^\star(g)\|}{\sqrt{N} J(p, \|g\|_{\sup}) + 1} \ge C\sqrt{18\log(Q_N)} \middle| \mathcal{F}_1^T \right) \\
\le \;& 2(Q_N + 2)\exp\left( -\frac{18C^2 \log(Q_N)}{C^2 \exp(2)} \right) \le \frac{2(Q_N + 2)}{Q_N^2}. \qquad\qquad (\text{S.12})
\end{aligned}
$$

Taking expectation on both sides of (S.12), we get that

$$P\left(\sup_{g\in\mathcal{G}(p)}\frac{\sqrt{N}\|Z_M^\star(g)\|}{\sqrt{N}J(p,\|g\|_{\sup})+1}\geq C\sqrt{18\log(Q_N)}\right)=o(1),\ \text{as } N\to\infty.$$

This shows that, with probability approaching one,

$$\sup_{g\in\mathcal{G}(p)}\frac{\sqrt{N}\|Z_M^\star(g)\|}{\sqrt{N}J(p,\|g\|_{\sup})+1}\leq C\sqrt{18\log(Q_N)}.$$

Since $\|g\|_{\sup}\leq 1$ for any $g\in\mathcal{G}$ and $J(p,\delta)$ is increasing in $\delta$, the above inequality implies that, with probability approaching one,

$$\sup_{g\in\mathcal{G}(p)}\|Z_M^\star(g)\|\leq C\sqrt{18\log(Q_N)}(J(p,1)+N^{-1/2}).$$

Combining with Lemma S.3, we get that

$$\begin{aligned}\sup_{g\in\mathcal{G}(p)}\|Z_M(g)\| &\leq \sup_{g\in\mathcal{G}(p)}\|Z_M(g)-Z_M^\star(g)\|+\sup_{g\in\mathcal{G}(p)}\|Z_M^\star(g)\|\\ &= O_P\left(1+\sqrt{\log\log\left(NJ(p,1)\right)}(J(p,1)+N^{-1/2})\right).\end{aligned}$$

Proof completed. □

*Proof of Lemma A.5.* By (2.5), we have $e_i'=\epsilon_i'-\Delta_i, \bar{v}=\epsilon_i'-\Delta_i(\bar{X}-\bar{\Gamma}_1'F_1'-\bar{\Gamma}_2'F_2')$ and

$$\begin{aligned}(Y_i-\tau_i g_\eta)'PK_{\mathbb{X}_i}&=[\tau_i(g_0-g_\eta)+\Sigma'\beta_i+e_i]'PK_{\mathbb{X}_i}\\ &=[\tau_i(g_0-g_\eta)+e_i]'PK_{\mathbb{X}_i}\\ &=[\tau_i(g_0-g_\eta)+\epsilon_i+F_2\bar{\Gamma}_2\Delta_i']'PK_{\mathbb{X}_i}.\end{aligned}\tag{S.13}$$

By the definition of $g_\eta$ in the proof Theorem 4.1 and (S.13), we get that

$$S_{M,\eta}(g_\eta)=S_{M,\eta}(g_\eta)-S_{M,\eta}^\star(g_\eta)=T_1+T_2-T_3+W_\eta g_\eta-E(W_\eta g_\eta|\mathcal{F}_1^T)=T_1+T_2-T_3,\tag{S.14}$$

where

$$\begin{aligned}T_1&=\frac{1}{NT}\sum_{i=1}^N[\epsilon_i'PK_{\mathbb{X}_i}-E(\epsilon_i'PK_{\mathbb{X}_i}|\mathcal{F}_1^T)],\\ T_2&=\frac{1}{NT}\sum_{i=1}^N[\Delta_iF_2'PK_{\mathbb{X}_i}-E(\Delta_iF_2'PK_{\mathbb{X}_i}|\mathcal{F}_1^T)],\\ T_3&=\kappa(g_\eta-g_0).\end{aligned}$$

Recall that $\kappa$ is defined in the proof of Theorem 4.1. It is worthwhile to mention that the terms $W_\eta g_\eta$ and $E(W_\eta g_\eta|\mathcal{F}_1^T)$ cancel each other in (S.14) thanks to $W_\eta g_\eta\in\mathcal{F}_1^T$. Next, we will bound $T_1, T_2, T_3$ respectively.

First of all, by (A.23) and (A.24), it yields that

$$\|T_3\| = o_P(1)\|g_\eta - g_0\| = o_P(\frac{1}{\sqrt{NTh}} + \frac{1}{N\sqrt{h}} + \sqrt{\eta}). \tag{S.15}$$

Secondly, the independence of $\epsilon_i$ and $\mathbb{X}_i, F_1, F_2$ tells us that

$$T_1 = \frac{1}{NT}\sum_{i=1}^{N} \epsilon_i' P K_{\mathbb{X}_i}.$$

Again by the independence assumption and direct calculations, we have

$$\begin{aligned}
E(\|T_1\|^2|\mathcal{F}_1^T) &= \frac{1}{N^2T^2}\sum_{i=1}^{N} E(\epsilon_i' P < K_{\mathbb{X}_i}, K_{\mathbb{X}_i} > P'\epsilon_i|\mathcal{F}_1^T) \\
&= \frac{1}{N^2T^2}\sum_{i=1}^{N} E(\epsilon_i' P \mathbf{K}_i P'\epsilon_i|\mathcal{F}_1^T) \\
&= \frac{1}{N^2T^2}\sum_{i=1}^{N} Tr(E(P\mathbf{K}_i P'\epsilon_i\epsilon_i')|\mathcal{F}_1^T) \\
&= \frac{\sigma_\epsilon^2}{NT^2} E\{Tr(P\mathbf{K}_i P')|\mathcal{F}_1^T\} \\
&\leq \frac{\sigma_\epsilon^2}{NT^2} E\{Tr(\mathbf{K}_i)|\mathcal{F}_1^T\} \\
&= O_P(\frac{1}{NTh}),
\end{aligned}$$

where we are using the facts that $\mathbf{K}_i = [K(X_{it}, X_{il})]_{1\leq t,l\leq T}$ and $Tr(\mathbf{K}_i) \leq Tc_\varphi^2 h^{-1}$ derived from (A.19). So it follows

$$\|T_1\| = O_P(\frac{1}{\sqrt{NTh}}) \tag{S.16}$$

Lastly, we will handle $T_2$ as follows. Since $F_2' P_\star = 0$ (see Section A.3), it follows that

$$T_2 = \frac{1}{NT}\sum_{i=1}^{N}[\Delta_i F_2'(P - P_\star)K_{\mathbb{X}_i} - E(\Delta_i F_2' P K_{\mathbb{X}_i}|\mathcal{F}_1^T)].$$

By the proof and notation in Lemma A.3, it can be shown that

$$\begin{aligned}
& P_\star - P \\
= \;& \Sigma'\left((\Sigma\Sigma')^{-1} - (\Sigma_\star\Sigma_\star')^{-1}\right)\Sigma + \Sigma_\star'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma} + \widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\Sigma_\star + \widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma}.
\end{aligned}$$

Consequently, $T_2$ has following decomposition:

$$\begin{aligned}
& T_2 \\
= & \frac{1}{NT}\sum_{i=1}^{N}[\Delta_i F_2'(\Sigma'\left((\Sigma\Sigma')^{-1}-(\Sigma_\star\Sigma_\star')^{-1}\right)\Sigma)K_{\mathbb{X}_i}-E(\Delta_i F_2'(\Sigma'\left((\Sigma\Sigma')^{-1}-(\Sigma_\star\Sigma_\star')^{-1}\right)\Sigma)K_{\mathbb{X}_i}|\mathcal{F}_1^T)] \\
& +\frac{1}{NT}\sum_{i=1}^{N}[\Delta_i F_2'\Sigma_\star'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma}K_{\mathbb{X}_i}-E(\Delta_i F_2'\Sigma_\star'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma}K_{\mathbb{X}_i}|\mathcal{F}_1^T)] \\
& +\frac{1}{NT}\sum_{i=1}^{N}[\Delta_i F_2'\widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\Sigma_\star K_{\mathbb{X}_i}-E(\Delta_i F_2'\widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\Sigma_\star K_{\mathbb{X}_i}|\mathcal{F}_1^T)] \\
& +\frac{1}{NT}\sum_{i=1}^{N}[\Delta_i F_2'\widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma}K_{\mathbb{X}_i}-E(\Delta_i F_2'\widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma}K_{\mathbb{X}_i}|\mathcal{F}_1^T)] \\
\equiv & T_{21}+T_{22}+T_{23}+T_{24}.
\end{aligned}$$

The rest of the proof proceeds to bound the terms $T_{2i}, i=1,2,3,4$. By (S.9) in the proof of Lemma A.3, we obtain the following:

$$\begin{aligned}
E(\|F_2'\Sigma'\left((\Sigma\Sigma')^{-1}-(\Sigma_\star\Sigma_\star')^{-1}\right)\Sigma\|_{\mathrm{op}}) &= O(N^{-1/2}+T^{1/2}/N), \\
E(\|F_2'\widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\Sigma_\star\|_{\mathrm{op}}) &= O(N^{-1/2}), \\
E(\|F_2'\widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma}\|_{\mathrm{op}}) &= O(1/N).
\end{aligned}$$

Therefore, it follows that

$$\begin{aligned}
& \|E(\frac{1}{NT}\sum_{i=1}^{N}\Delta_i F_2'\Sigma'\left((\Sigma\Sigma')^{-1}-(\Sigma_\star\Sigma_\star')^{-1})\right)\Sigma K_{\mathbb{X}_i}|\mathcal{F}_1^T)\| \\
\le & \frac{1}{NT}\sum_{i=1}^{N}E(\|\Delta_i F_2'\Sigma'\left((\Sigma\Sigma')^{-1}-(\Sigma_\star\Sigma_\star')^{-1})\right)\Sigma K_{\mathbb{X}_i}\||\mathcal{F}_1^T) \\
\le & \frac{1}{NT}\sum_{i=1}^{N}\|\Delta_i\|_2 E\left(\|F_2'\Sigma'\left((\Sigma\Sigma')^{-1}-(\Sigma_\star\Sigma_\star')^{-1})\right)\Sigma\|_{\mathrm{op}}\sqrt{\sum_{t=1}^{T}\|K_{X_{it}}\|^2}|\mathcal{F}_1^T\right) \\
\le & \frac{1}{NT}\sum_{i=1}^{N}\|\Delta_i\|_2 E\left(\|F_2'\Sigma'\left((\Sigma\Sigma')^{-1}-(\Sigma_\star\Sigma_\star')^{-1})\right)\Sigma\|_{\mathrm{op}}\sqrt{Tc_\varphi^2 h^{-1}}|\mathcal{F}_1^T\right) \\
\le & \frac{1}{\sqrt{T}}\sup_{1\le i\le N}\|\Delta_i\|_2\sqrt{c_\varphi^2 h^{-1}}E(\|F'\Sigma'\left((\Sigma\Sigma')^{-1}-(\Sigma_\star\Sigma_\star')^{-1}\right)\Sigma\|_{\mathrm{op}}|\mathcal{F}_1^T) \\
= & O_P(\frac{1}{\sqrt{NTh}}+\frac{1}{N\sqrt{h}}).
\end{aligned}$$

As a consequence, $\|T_{21}\| = O_P((NTh)^{-1/2} + N^{-1}h^{-1/2})$. Similarly,

$$E\{\|E(\frac{1}{NT}\sum_{i=1}^{N}\Delta_i F_2'\widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\Sigma_\star K_{\mathbb{X}_i}|\mathcal{F}_1^T)\|\} = O_P(\frac{1}{\sqrt{NTh}}),$$
$$E\{\|E(\frac{1}{NT}\sum_{i=1}^{N}\Delta_i F_2'\widetilde{\Sigma}'(\Sigma_\star\Sigma_\star')^{-1}\widetilde{\Sigma} K_{\mathbb{X}_i}|\mathcal{F}_1^T)\|\} = O_P(\frac{1}{N\sqrt{Th}}).$$

So it follows that $\|T_{23}\| = O_P((NTh)^{-1/2})$ and $\|T_{24}\| = O_P(N^{-1}(Th)^{-1/2})$. Finally, we will handle $T_{22}$. Let $W = F_2'\Sigma_\star'(\Sigma_\star\Sigma_\star')^{-1}$. It can be easily seen from (S.10) that $W \in \mathcal{F}_1^T$ and $\|W\|_{\text{op}} = O_P(1)$. To bound $T_{22}$, notice

$$\widetilde{\Sigma}K_{\mathbb{X}_i} = \begin{pmatrix} 0_{q_1\times T} \\ \sum_{t=1}^T \bar{v}_t K_{X_{it}} \end{pmatrix} = \begin{pmatrix} 0_{q_1\times T} \\ \sum_{t=1}^T \bar{v}_{t1} K_{X_{it}} \\ \sum_{t=1}^T \bar{v}_{t2} K_{X_{it}} \\ \cdots \\ \sum_{t=1}^T \bar{v}_{td} K_{X_{it}} \end{pmatrix},$$

where $\bar{v}_{ti}$ is the $i$th element of vector $\bar{v}_t$. By direct calculations, it follows that

$$\begin{aligned} \|T_{22}\| &= \|\frac{1}{NT}\sum_{i=1}^{N}\{\Delta_i W\widetilde{\Sigma}K_{\mathbb{X}_i} - E(\Delta_i W\widetilde{\Sigma}K_{\mathbb{X}_i}|\mathcal{F}_1^T)\}\| \\ &\le \frac{1}{NT}\sum_{i=1}^{N}\|\Delta_i W\widetilde{\Sigma}K_{\mathbb{X}_i} - E(\Delta_i W\widetilde{\Sigma}K_{\mathbb{X}_i}|\mathcal{F}_1^T)\| \\ &\le \frac{1}{NT}\sum_{i=1}^{N}\|\Delta_i\|_2\|W\|_{\text{op}}\sqrt{\sum_{l=1}^{d}\|\sum_{t=1}^{T}\left(\bar{v}_{tl}K_{X_{it}} - E(\bar{v}_{tl}K_{X_{it}}|\mathcal{F}_1^T)\right)\|^2}. \end{aligned} \tag{S.17}$$

By (S.17), it suffices to find the rate of

$$\frac{1}{NT}\sum_{i=1}^{N}\sqrt{\sum_{l=1}^{d}\|\sum_{t=1}^{T}\left(\bar{v}_{tl}K_{X_{it}} - E(\bar{v}_{tl}K_{X_{it}}|\mathcal{F}_1^T)\right)\|^2}. \tag{S.18}$$

Because $d$ is finite and fixed, to simplify our technical arguments, assume $d = 1$ without loss of generality. Direct examinations give the following decomposition:

$$\begin{aligned} &\|\sum_{t=1}^{T}\left(\bar{v}_{tl}K_{X_{it}} - E(\bar{v}_{tl}K_{X_{it}}|\mathcal{F}_1^T)\right)\|^2 \\ =\ & \frac{1}{N^2}\|\sum_{t=1}^{T}\sum_{j=1}^{N}\left(v_{jtl}K_{X_{it}} - E(v_{jtl}K_{X_{it}}|\mathcal{F}_1^T)\right)\|^2 \\ =\ & \frac{2}{N^2}\|\sum_{t=1}^{T}\sum_{j\neq i}\left(v_{jtl}K_{X_{it}} - E(v_{jtl}K_{X_{it}}|\mathcal{F}_1^T)\right)\|^2 + \frac{2}{N^2}\|\sum_{t=1}^{T}\left(v_{itl}K_{X_{it}} - E(v_{itl}K_{X_{it}}|\mathcal{F}_1^T)\right)\|^2 \\ \equiv\ & T_{221} + T_{222}. \end{aligned}$$

When $i \neq j$, $v_{jtl}$ is independent of $X_{it}, \mathcal{F}_1^T$, so it follows that

$$
\begin{aligned}
& E\{\| \sum_{t=1}^{T} \sum_{j\neq i} \left(v_{jtl} K_{X_{it}} - E(v_{jtl} K_{X_{it}} | \mathcal{F}_1^T)\right) \|^2 | \mathcal{F}_1^T\} \\
= \ & E\{\| \sum_{t=1}^{T} \sum_{j\neq i} v_{jtl} K_{X_{it}} \|^2 | \mathcal{F}_1^T\} \\
= \ & E\{ \sum_{t,t'=1}^{T} \sum_{j,j'\neq i} v_{jtl} v_{j't'l} K(X_{it}, X_{it'}) | \mathcal{F}_1^T\} \\
= \ & \sum_{t=1}^{T} \sum_{j\neq i} E(v_{jtl}^2) E(K(X_{it}, X_{it})\} | \mathcal{F}_1^T) \\
\leq \ & NTE(v_{11l}^2) c_\varphi^2 h^{-1},
\end{aligned}
$$

As a consequence, $T_{221} = O_P(T(Nh)^{-1})$. To deal with $T_{222}$, by Cauchy inequality, it yields that

$$
\begin{aligned}
E\{\| \sum_{t=1}^{T} (v_{itl} K_{X_{it}}) \|^2 | \mathcal{F}_1^T\} & \leq E\{ \sum_{t=1}^{T} v_{itl}^2 \sum_{t=1}^{T} \|K_{X_{it}}\|^2 | \mathcal{F}_1^T\} \\
& \leq E(\sum_{t=1}^{T} v_{itl}^2) T c_\varphi^2 h^{-1} \\
& = E(v_{11l}^2) T^2 c_\varphi h^{-1},
\end{aligned}
$$

which further implies $T_{222} = O_P(T^2(N^2h)^{-1})$. By Jensen's inequality and $d = 1$, it follows that

$$
\begin{aligned}
(S.18) & = E\left( \frac{1}{NT} \sum_{i=1}^{N} \sqrt{ \| \sum_{t=1}^{T} \left( \bar{v}_{tl} K_{X_{it}} - E(\bar{v}_{tl} K_{X_{it}} | \mathcal{F}_1^T) \right) \|^2 | \mathcal{F}_1^T } \right) \\
& \leq \frac{1}{NT} \sum_{i=1}^{N} \sqrt{ E\left( \| \sum_{t=1}^{T} \left( \bar{v}_{tl} K_{X_{it}} - E(\bar{v}_{tl} K_{X_{it}} | \mathcal{F}_1^T) \right) \|^2 | \mathcal{F}_1^T \right) } \\
& \leq \sqrt{ 2c_\varphi^2 E(v_{11l}^2) (\frac{1}{NTh} + \frac{1}{N^2 h}) } \\
& = O_P(\frac{1}{\sqrt{NTh}} + \frac{1}{N\sqrt{h}}). \qquad (S.19)
\end{aligned}
$$

Combining (S.17) and (S.19), it follows that $\|T_{22}\| = O_P((NTh)^{-1/2} + (Nh^{1/2})^{-1})$. As a consequence, we have

$$
\|T_2\| = O_P(\frac{1}{\sqrt{NTh}} + \frac{1}{N\sqrt{h}}). \qquad (S.20)
$$

Combining (S.15), (S.16) and (S.20), it yields that

$$
\|S_{M,\eta}(g_\eta)\| = O_P(\frac{1}{\sqrt{NTh}} + \frac{1}{N\sqrt{h}}) + o_P(\sqrt{\eta}).
$$

Proof completed. □

Next we will prove Lemmas A.6, A.7 and A.8. For this purpose, let us introduce a set of notation. Define $V_{NT\star}, A_{NT\star}, V_{NTm\star}, A_{NTm\star}, H_{NTm\star}$ as follows,

$$V_{NT\star} = \frac{1}{NT}\sum_{i=1}^{N} K_{\mathbb{X}_i}(x_0)' P_\star K_{\mathbb{X}_i}(x_0), A_{NT\star} = V_{NT\star}^{-1/2},$$

$$V_{NTm\star} = \frac{1}{NT}\sum_{i=1}^{N} \phi_m' \Phi_i' P_\star \Phi_i \phi_m, A_{NTm\star} = V_{NTm\star}^{-1/2}, H_{NTm\star} = \frac{1}{NT}\sum_{i=1}^{N} \Phi_i' P_\star \Phi_i.$$

*Proof of Lemma A.6.* Define

$$Q_{i\star} = E(\frac{\Phi_i' P_\star \Phi_i}{T}|\mathcal{F}_1^T), \bar{Q}_\star = \frac{1}{N}\sum_{i=1}^{N} Q_{i\star}, Q_i = E(\frac{\Phi_i' P \Phi_i}{T}|\mathcal{F}_1^T), \bar{Q} = \frac{1}{N}\sum_{i=1}^{N} Q_i = I_m.$$

Notice that, conditioning on $\mathcal{F}_1^T$, $\Phi_i$ are independent. Hence, by Chebyshev's inequality, it follows that

$$\begin{aligned}
P(\|H_{NTm\star} - \bar{Q}_\star\|_F > \epsilon|\mathcal{F}_1^T) &= P(\|\frac{1}{N}\sum_{i=1}^{N}(\frac{\Phi_i' P_\star \Phi_i}{T} - Q_{i\star})\|_F > \epsilon|\mathcal{F}_1^T) \\
&\leq \frac{1}{\epsilon^2 N^2} E\{Tr\left([\sum_{i=1}^{N}(\frac{\Phi_i' P_\star \Phi_i}{T} - Q_{i\star})]^2\right)|\mathcal{F}_1^T\} \\
&= \frac{1}{\epsilon^2 N^2}\sum_{i=1}^{N} Tr\{E\left([\frac{\Phi_i' P_\star \Phi_i}{T} - Q_{i\star}]^2\right)|\mathcal{F}_1^T\} \\
&= \frac{1}{\epsilon^2 N^2}\sum_{i=1}^{N} E\left(\|\frac{\Phi_i' P_\star \Phi_i}{T} - Q_{i\star}\|_F^2|\mathcal{F}_1^T\right) \\
&\leq \frac{1}{\epsilon^2 N^2}\sum_{i=1}^{N} E\left(\|\frac{\Phi_i' P_\star \Phi_i}{T}\|_F^2|\mathcal{F}_1^T\right) \\
&\leq \frac{1}{\epsilon^2 N^2 T^2}\sum_{i=1}^{N} E\left(\|\Phi_i\|_F^4|\mathcal{F}_1^T\right) \\
&\leq \frac{m^2(c_\varphi+1)^4}{\epsilon^2 N}.
\end{aligned}$$

As a consequence, it follows that,

$$P(\|H_{NTm\star} - \bar{Q}_\star\|_F > \frac{\epsilon m(c_\varphi+1)^2}{\sqrt{N}}|\mathcal{F}_1^T) \leq \frac{1}{\epsilon^2}.$$

By taking expectation on both sides, we have

$$P(\|H_{NTm\star} - \bar{Q}_\star\|_F > \frac{\epsilon m(c_\varphi+1)^2}{\sqrt{N}}) \leq \frac{1}{\epsilon^2}.$$

Since $c_\varphi = O_P(1)$, we obtain

$$\|H_{NTm\star} - \bar{Q}_\star\|_F = O_P(mN^{-1/2}). \tag{S.21}$$

By Lemma A.3, we have

$$\begin{aligned} E(\|H_{NTm} - H_{NTm\star}\|_F|\mathcal{F}_1^T) &\leq \frac{1}{NT}\sum_{i=1}^N E(\|\Phi_i'(P-P_\star)\Phi_i\|_F|\mathcal{F}_1^T) \\ &\leq \frac{1}{NT}\sum_{i=1}^N E(\|(P-P_\star)\|_{\text{op}}\|\Phi_i\|_F^2|\mathcal{F}_1^T) \\ &\leq \frac{1}{NT}\sum_{i=1}^N mTc_\varphi^2 E(\|(P-P_\star)\|_{\text{op}}|\mathcal{F}_1^T) \\ &= O_P(\frac{m}{\sqrt{N}}). \end{aligned} \tag{S.22}$$

Again by Lemma A.3 and similar calculations, it follows that

$$\begin{aligned} \|\bar{Q} - \bar{Q}_\star\|_F &= \|E(\frac{1}{NT}\sum_{i=1}^N \Phi_i'(P-P_\star)\Phi_i|\mathcal{F}_1^T)\|_F \\ &\leq \frac{1}{NT}\sum_{i=1}^N E(\|\Phi_i'(P-P_\star)\Phi_i\|_F|\mathcal{F}_1^T) \\ &= O_P(\frac{m}{\sqrt{N}}). \end{aligned} \tag{S.23}$$

Combining (S.21), (S.22) and (S.23), it yields that

$$\|H_{NTm} - I_m\|_F = \|H_{NTm} - \bar{Q}\|_F = O_P(\frac{m}{\sqrt{N}}) = o_P(1).$$

To the end of the proof, we quantify the minimal and maximal eigenvalues of $H_{NTm}$ as follows.

$$\begin{aligned} \lambda_{\min}(H_{NTm}) &= \min_{\|u\|_2=1} u'H_{NTm}u \\ &\geq \min_{\|u\|_2=1} u'I_m u - \min_{\|u\|_2=1} |u'(H_{NTm} - I_m)u| \\ &= 1 - \|H_{NTm} - I_m\|_{\text{op}} \\ &= 1 + o_P(1), \end{aligned}$$

and

$$\begin{aligned} \lambda_{\max}(H_{NTm}) &= \max_{\|u\|_2=1} u'H_{NTm}u \\ &\leq \max_{\|u\|_2=1} u'I_m u + \max_{\|u\|_2=1} |u'(H_{NTm} - I_m)u| \\ &= 1 + \|H_{NTm} - I_m\|_{\text{op}} \\ &= 1 + o_P(1), \end{aligned}$$

where have used the trivial inequality $\|H_{NTm} - I_m\|_{\text{op}} \leq \|H_{NTm} - I_m\|_F = o_P(1)$. Proof completed.

□

*Proof of Lemma A.7.* By Lemma A.6, we find a lower bound for $V_{NTm}$ and an upper bound for $A_{NTm}$ as follows:

$$\begin{aligned} V_{NTm} &= \phi_m' H_{NTm}\phi_m \geq \lambda_{\min}(H_{NTm})\|\phi_m\|_2^2 \geq \lambda_{\min}(H_{NTm})C, \\ A_{NTm} &= V_{NTm}^{-1/2} \leq \lambda_{\min}^{-1/2}(H_{NTm})\|\phi_m'\|_2^{-1} = O_P(1)(\sum_{\nu=1}^{m}\frac{\varphi_\nu^2(x_0)}{(1+\eta\rho_\nu)^2})^{-1/2} = O_P(1). \quad \text{(S.24)} \end{aligned}$$

Define $L_i(x_0) = K_{\mathbb{X}_i}(x_0) - \Phi_i\phi_m$. Then it follows that

$$E(\|L_i\|_2^2|\mathcal{F}_1^T) \leq Tc_\varphi^4(\sum_{\nu=m+1}^{\infty}\frac{1}{1+\eta\rho_\nu})^2 \equiv Tc_\varphi^4 D_m^2.$$

Directly calculation shows that

$$|V_{NT} - V_{NTm}| \leq |\frac{2}{NT}\sum_{i=1}^{N} L_i' P K_{\mathbb{X}_i}| + |\frac{1}{NT}\sum_{i=1}^{N} L_i' P L_i| \equiv 2|T1| + |T2|. \quad \text{(S.25)}$$

Let $R_{x_0}(\cdot) = \sum_{\nu=m+1}^{\infty}\frac{\varphi_\nu(x_0)\varphi_\nu(\cdot)}{1+\eta\rho_\nu}$. Notice $L_i = \tau_i R_{x_0}$ and $E(T_1|\mathcal{F}_1^T) = V(K_{x_0}, R_{x_0})$. Similar to the proof of Lemma A.6, we can show that

$$E(|T_1 - V(K_{x_0}, R_{x_0})||\mathcal{F}_1^T) \quad = \quad O_P(\frac{D_m}{\sqrt{N}h}).$$

Meanwhile we have the following

$$V(K_{x_0}, R_{x_0}) = \sum_{\nu=m+1}^{\infty}\frac{\varphi_\nu^2(x_0)}{(1+\eta\rho_\nu)^2} \leq c_\varphi^2 D_m.$$

As a consequence, it follows that $|T_1| = O_P(D_m)$.

A bound for $T_2$ is given by the following inequality,

$$E(|T_2||\mathcal{F}_1^T) \leq \frac{1}{NT}\sum_{i=1}^{N} E(\|L_i\|_2^2|\mathcal{F}_1^T) = O_P(D_m^2).$$

So (S.25) becomes $V_{NT} - V_{NTm} = O_P(D_m) = o_P(1)$. Hence $A_{NT} = A_{NTm} + o_P(1) = O_P(1)$, where last equality is from (S.24). Proof completed. □

*Proof of Lemma A.8.* The proof of this lemma is based on Lyapunov C.L.T. Let $c_i = A_{NTm}P\Phi_i\phi_m/(NT)$. We have

$$\sqrt{NT}A_{NTm}(\frac{1}{NT}\sum_{i=1}^{N}\phi_m'\Phi_i'P\epsilon_i) = \sum_{i=1}^{N}\sqrt{NT}c_i'\epsilon_i.$$

Since $c_i \in \mathcal{D}_1^T$ and $\epsilon_i$ is independent of $\mathcal{D}_1^T$, it follows that

$$\begin{aligned} E[(\sum_{i=1}^{N}\sqrt{NT}c_i'\epsilon_i)^2||\mathcal{D}_1^T] &= NT\sigma_\epsilon^2\sum_{i=1}^{N}c_i'c_i \\ &= NT\sigma_\epsilon^2 A_{NTm}^2\frac{1}{N^2T^2}\sum_{i=1}^{N}\phi_m'\Phi_i'P\Phi_i\phi_m \\ &= \sigma_\epsilon^2. \end{aligned}$$

Let $c_{it}$ be the $t$th element of $c_i$. By direct examinations, it follows that

$$\begin{aligned}\sum_{i=1}^{N} E[(\sqrt{NT}c_i'\epsilon_i)^4||\mathcal{D}_1^T] &= N^2T^2\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it}^4 E(\epsilon_{it}^4) \\ &\quad +3N^2T^2\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{t'\neq t} c_{it}^2 c_{it'}^2 E(\epsilon_{it}^2\epsilon_{it'}^2). \qquad (S.26)\end{aligned}$$

Next we are going to find a bound for $c_{it}$. By direct calculation, we have

$$\begin{aligned}|c_{it}| &= |A_{NTm}\frac{1}{NT}p_{t\cdot}\Phi_i\phi_m| \\ &\leq \|A_{NTm}\phi_m\|_2\|\frac{1}{NT}\sum_{s=1}^{T} p_{ts}\Phi_{i,s\cdot}\|_2 \\ &\leq \lambda_{\min}^{-1/2}(H_{NTm})\frac{1}{NT}\|\sum_{s=1}^{T} p_{ts}\Phi_{i,s\cdot}\|_2,\end{aligned}$$

where $p_{t\cdot}$ is the $t$th row of $P$, $p_{ts}$ is the $(t,s)$th element of $P$ and $\Phi_{i,s\cdot}$ is the $s$th row of $\Phi_i$. Meanwhile, $p_{ts} = \delta_{ts} - Z_s'(\Sigma\Sigma')^{-1}Z_t$, hence

$$\begin{aligned}\|\sum_{s=1}^{T} p_{ts}\Phi_{i,s\cdot}\|_2 &= \|\Phi_{i,t\cdot} - \frac{1}{T}Z_t'(\frac{\Sigma\Sigma'}{T})^{-1}\sum_{s=1}^{T} Z_s\Phi_{i,s\cdot}\|_2 \\ &\leq \|\Phi_{i,t\cdot}\|_2 + \|Z_t\|_2\|(\frac{\Sigma\Sigma'}{T})^{-1}\|_{\text{op}}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\|Z_s\|_2^2}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\|\Phi_{i,s\cdot}\|_2^2} \\ &\leq \sqrt{mc_\varphi^2} + \|Z_t\|_2\|(\frac{\Sigma\Sigma'}{T})^{-1}\|_{\text{op}}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\|Z_s\|_2^2}\sqrt{mc_\varphi^2} \\ &\leq \sqrt{mc_\varphi^2}(1 + b\|Z_t\|_2),\end{aligned}$$

where $b = \|(\frac{\Sigma\Sigma'}{T})^{-1}\|_{\text{op}}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\|Z_s\|_2^2} = O_P(1)$ by Assumption A5. So $|c_{it}| \leq a(1 + b\|Z_t\|_2)$, where $a = \lambda_{\min}^{-1/2}(H_{NTm})\frac{1}{NT}\sqrt{mc_\varphi^2}$. By Lemma A.6, we have

$$\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it}^4 \leq \sum_{i=1}^{N}\sum_{t=1}^{T} 8a^4(1 + b^4\|Z_t\|_2^4) = O_P(\frac{m^2}{N^3T^3}), \qquad (S.27)$$

and

$$\begin{aligned}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{t'\neq t} c_{it}^2c_{it'}^2 &\leq \sum_{i=1}^{N}\sqrt{\sum_{t=1}^{T}\sum_{t'\neq t} c_{it}^4}\sqrt{\sum_{t=1}^{T}\sum_{t'\neq t} c_{it'}^4} \\ &\leq T\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it}^4 \\ &= O_P(\frac{m^2}{N^3T^2}). \qquad (S.28)\end{aligned}$$

Combining (S.26), (S.27) and (S.28), we have $\sum_{i=1}^{N} E[(\sqrt{NT} c_i' \epsilon_i)^4 || \mathcal{D}_1^T] = O_P(m^2/N)$. And by Lyapunov C.L.T, the result follows. Proof completed. □