Supplemental Appendix to
Uniform Nonparametric Inference for Time Series

Jia Li* and Zhipeng Liao†

April 25, 2020

Abstract
Supplemental Appendix S.A contains the proofs for all results in the main text. Supplemental Appendix S.B contains additional technical results on the verification of high-level conditions using more primitive ones.

*Department of Economics, Duke University, Durham, NC 27708; e-mail: jl410@duke.edu.
†Department of Economics, UCLA, Los Angeles, CA 90095; e-mail: zhipeng.liao@econ.ucla.edu.
S.A  Appendix: Proofs

For any real matrix $A$, we use $\|A\|$ and $\|A\|_S$ to denote its Frobenius norm and spectral norm, respectively. If $A$ is a real square matrix, we denote its trace, the smallest and the largest eigenvalues by $\text{Tr}(A)$, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$, respectively. We use $a^{(j)}$ to denote the $j$th component of a vector $a$; $A^{(i,j)}$ is defined similarly for a matrix $A$. For a random matrix $X$, $\|X\|_p$ denotes its $L_p$-norm, that is, $\|X\|_p = (\mathbb{E} \|X\|^p)^{1/p}$. For any two positive sequences $a_n$ and $b_n$, $a_n \ll b_n$ means that $a_n = o(b_n)$. For any two real constants $a$ and $b$, $a \land b = \min\{a, b\}$. Throughout the proofs, we use $K$ to denote a generic constant that may change from line to line.

S.A.1 Proof of Theorem 1

The proof of Theorem 1 consists of two steps. The first step is to approximate $S_n$ with another martingale $S_n^*$ whose conditional covariance matrix is exactly $\Sigma_n$; see Lemma A1. We then establish the coupling between $S_n^*$ and $\tilde{S}_n$ by using Lindeberg’s method and Strassen’s theorem; see Lemma A2.

Turning to the details, we start with describing the approximating martingale $S_n^*$. Consider the following stopping time:

$$
\tau_n \equiv \max \left\{ t \in \{1, \ldots, k_n\} : \Sigma_n - \sum_{s=1}^{t} V_{n,s} \text{ is positive semi-definite} \right\},
$$

with the convention that $\max \emptyset = 0$. We note that $\tau_n$ is a stopping time because $V_{n,t}$ is $\mathcal{F}_{n,t-1}$-measurable for each $t$ and $\Sigma_n$ is nonrandom. The matrix

$$
\xi_n \equiv \begin{cases} 
\Sigma_n & \text{when } \tau_n = 0, \\
\Sigma_n - \sum_{t=1}^{\tau_n} V_{n,t} & \text{when } \tau_n \geq 1,
\end{cases}
$$

is positive semi-definite by construction.

Let $K_n$ be a sequence of integers such that $K_n \to \infty$ and let $(\eta_{n,t})_{k_n+1 \leq t \leq k_n+K_n}$ be independent $m_n$-dimensional standard normal vectors. We construct another martingale difference array $(Z_{n,t}, \mathcal{H}_{n,t})_{1 \leq t \leq k_n+K_n}$ as follows:

$$
Z_{n,t} \equiv \begin{cases} 
X_{n,t} 1_{\{t \leq \tau_n\}} & \text{when } 1 \leq t \leq k_n, \\
K_n^{-1/2} \xi_n^{1/2} \eta_{n,t} & \text{when } k_n + 1 \leq t \leq k_n + K_n,
\end{cases}
$$

and the filtration is given by

$$
\mathcal{H}_{n,t} \equiv \begin{cases} 
\mathcal{F}_{n,t} & \text{when } 1 \leq t \leq k_n, \\
\mathcal{F}_{n,k_n} \lor \sigma(\eta_{n,s} : s \leq t) & \text{when } k_n + 1 \leq t \leq k_n + K_n.
\end{cases}
$$

Since $\tau_n$ is a stopping time, it is easy to verify that $(Z_{n,t}, \mathcal{H}_{n,t})_{1 \leq t \leq k_n+K_n}$ indeed forms a martingale difference array. We denote

$$
V_{n,t}^* = \mathbb{E} \left[ Z_{n,t} Z_{n,t}^\top \mathcal{H}_{n,t-1} \right] \tag{A.1}
$$
and set
\[ S^*_n = \sum_{t=1}^{k_n+K_n} Z_{n,t}. \] (A.2)

The conditional covariance matrix of \( S^*_n \) is exactly \( \Sigma_n \), that is,
\[ \sum_{t=1}^{k_n+K_n} V^*_n,t = \sum_{t=1}^{\tau_n} V_{n,t} + \xi_n = \Sigma_n. \] (A.3)

Lemma A1, below, quantifies the approximation error between \( S_n \) and \( S^*_n \).

**Lemma A1.** Suppose that Assumption 1 holds. Then, \( \| S_n - S^*_n \| = O_p(m_n^{1/2} r_n^{1/2}) \).

**Proof of Lemma A1.** Step 1. In this step, we show that for any \( \varepsilon > 0 \), there exists a finite constant \( C_1 > 0 \) such that, for \( u^*_n = \lceil C_1 r_n k_n \rceil \) and \( h^*_n = k_n - u^*_n \),
\[ \limsup_{n \to \infty} P(\tau_n < h^*_n) < \varepsilon. \] (A.4)

Fix \( \varepsilon > 0 \). By Assumption 1(ii), there exists a finite constant \( C_2 > 0 \) such that for any \( h_n \leq k_n \) satisfying \( h_n/k_n \to 1 \),
\[ \limsup_{n \to \infty} P(\lambda_{\text{max}}(\sum_{t=1}^{h_n} V_{n,t} - \Sigma_n h_n) > C_2 r_n) < \varepsilon. \] (A.5)

Let \( \Lambda > 0 \) denote a lower bound for the eigenvalues as described in Assumption 1(i). We shall show that (A.4) holds for \( C_1 \equiv C_2/\Lambda \).

Since \( r_n = o(1) \) by Assumption 1(ii), we have \( u^*_n/k_n \to 0 \) and \( h^*_n/k_n \to 1 \). In particular, (A.5) holds for \( h_n = h^*_n \). Moreover, observe that
\[ \frac{u^*_n}{r_n k_n} = \frac{\lceil C_1 r_n k_n \rceil}{r_n k_n} \geq \frac{C_1}{\Lambda}, \]
which, together with the definition of \( \Lambda \), implies that
\[ C_2 r_n \leq \frac{u^*_n}{k_n} \leq \lambda_{\text{min}} \left( \sum_{t=h^*_n+1}^{k_n} \mathbb{E}[V_{n,t}] \right). \] (A.6)

We then observe
\[
\begin{align*}
P(\tau_n < h^*_n) &\leq P \left( \lambda_{\text{max}} \left( \sum_{t=1}^{h^*_n} V_{n,t} - \Sigma_n \right) > 0 \right) \\
&= P \left( \lambda_{\text{max}} \left( \sum_{t=1}^{h^*_n} V_{n,t} - \Sigma_n h^*_n - (\Sigma_n - \Sigma_n h^*_n) \right) > 0 \right) \\
&\leq P \left( \lambda_{\text{max}} \left( \sum_{t=1}^{h^*_n} V_{n,t} - \Sigma_n h^*_n \right) > \lambda_{\text{min}} \left( \sum_{t=h^*_n+1}^{k_n} \mathbb{E}[V_{n,t}] \right) \right) \\
&\leq P \left( \lambda_{\text{max}} \left( \sum_{t=1}^{h^*_n} V_{n,t} - \Sigma_n h^*_n \right) > C_2 r_n \right),
\end{align*}
\] (A.7)
where the first inequality follows from the definition of $\tau_n$, the second inequality follows from the property of eigenvalues and the last inequality is by (A.6). From (A.5) and (A.7), the claim (A.4) readily follows.

**Step 2.** We now prove the assertion of Lemma A1. Note that

$$S_n - S_n^* = \sum_{t=1}^{k_n} X_{n,t} 1\{t > \tau_n\} - K_n^{-1/2} \xi_n^{1/2} \sum_{t=k_n+1}^{k_n + K_n} \eta_{n,t}.$$ 

Hence, it suffices to show

$$\sum_{t=1}^{k_n} X_{n,t} 1\{t > \tau_n\} = O_p(m_n^{1/2} r_n^{1/2}), \quad K_n^{-1/2} \xi_n^{1/2} \sum_{t=k_n+1}^{k_n + K_n} \eta_{n,t} = O_p(m_n^{1/2} r_n^{1/2}). \quad (A.8)$$

Recall $u_n^*$ and $h_n^*$ from step 1. By the assertion of step 1, we can assume that $\tau_n \geq h_n^*$ without loss of generality; otherwise, we can restrict attention to the event $\{\tau_n \geq h_n^*\}$ with the exceptional probability made arbitrarily small.

Since $\tau_n$ is a stopping time, $\{t > \tau_n\} \in F_{n,t-1}$. Therefore, $(X_{n,t} 1\{t > \tau_n\})_{t \geq 1}$ are martingale differences. It is then easy to see that

$$E \left[ \left\| \sum_{t=1}^{k_n} X_{n,t} 1\{t > \tau_n\} \right\|^2 \right] = E \left[ \sum_{t=1}^{k_n} \|X_{n,t}\|^2 1\{t > \tau_n\} \right] \leq \sum_{t=h_n^* + 1}^{k_n} E \left[ \|X_{n,t}\|^2 \right] = \text{Tr} \left[ \sum_{t=h_n^* + 1}^{k_n} E [V_{n,t}] \right].$$

By Assumption 1(i), the majorant side of the above inequality is $O((u_n^* m_n / k_n) = O(m_n r_n)$. The first assertion in (A.8) then readily follows.

Turning to the second assertion in (A.8), we note that

$$E \left[ \left\| K_n^{-1/2} \xi_n^{1/2} \sum_{t=k_n+1}^{k_n + K_n} \eta_{n,t} \right\|^2 \right] = \frac{1}{K_n} \sum_{t=k_n+1}^{k_n + K_n} E \left[ \|\xi_n^{1/2} \eta_{n,t}\|^2 \right] \leq \text{Tr} (E [\xi_n]) \leq \text{Tr} \left[ \sum_{t=h_n^* + 1}^{k_n} E [V_{n,t}] \right].$$

By the same argument as above, the majorant side of the above inequality is $O(m_n r_n)$, which implies the second assertion in (A.8). \(Q.E.D.\)

The next lemma establishes the strong approximation for $S_n^*$.

**Lemma A2.** Let $\bar{\lambda}$ denote the upper bound of the eigenvalues of $\Sigma_n$. Suppose that $K_n \geq 36 m_n^3 \bar{\lambda}^3 / B_n^2$. Then, there exists a sequence $\tilde{S}_n$ of $m_n$-dimensional centered Gaussian random vectors with covariance matrix $\Sigma_n$ such that

$$\left\| S_n^* - \tilde{S}_n \right\| = O_p((B_n m_n)^{1/3}).$$

4
Proof of Lemma A2. Step 1. We introduce some notations and outline the proof in this step. For any positive constant $C > 1$, we denote $\delta_{C,n} \equiv C(B_n m_n)^{1/3}$. We also set $\sigma_n^2 \equiv B_n^{2/3} m_n^{-1/3}$ and note that
\[
\frac{\delta_{C,n}^2}{m_n \sigma_n^2} = C^2 \text{ and } \frac{B_n}{\sigma_n^2 \delta_{C,n}} = C^{-1}.
\] (A.9)

Below, we denote
\[
\psi_{C,n} \equiv \left( \frac{C^2 \exp \left(C^2 - 1\right)}{m_n / 2} \right).
\]
Note that as $C \to \infty$,
\[
\psi_{C,n} \to 0 \text{ uniformly in } n.
\] (A.10)

In step 2, below, we show that the following inequality holds for any Borel subset $A \subseteq \mathbb{R}^{m_n}$:
\[
\mathbb{P} \left( S_n^* \in A \right) \leq F_n \left( A^{3\delta_{C,n}} \right) + \frac{1}{1 - \psi_{C,n}} \left( \psi_{C,n} + \frac{4B_n}{\sigma_n^2 \delta_{C,n}} \right),
\] (A.11)
where $F_n$ denotes the distribution of an $\mathcal{N}(0, \Sigma_n)$ random variable and
\[
A^{3\delta_{C,n}} \equiv \left\{ x \in \mathbb{R}^{m_n} : \inf_{y \in A} \|x - y\| \leq 3\delta_{C,n} \right\}.
\]

Consequently, by Strassen’s Theorem (see, e.g., Theorem 10.8 in Pollard (2001)), we can construct a variable $\tilde{S}_n \sim \mathcal{N}(0, \Sigma_n)$ such that
\[
\mathbb{P} \left( \|S_n^* - \tilde{S}_n\| > 3\delta_{C,n} \right) \leq \frac{1}{1 - \psi_{C,n}} \left( \psi_{C,n} + \frac{4B_n}{\sigma_n^2 \delta_{C,n}} \right) = \frac{1}{1 - \psi_{C,n}} \left( \psi_{C,n} + 4C^{-1} \right).
\]

By (A.10), for any $\varepsilon > 0$ there exists a constant $C_\varepsilon > 1$ such that for any $C > C_\varepsilon$ and for any $n$, the majorant side of the above inequality is bounded by $\varepsilon$, yielding
\[
\mathbb{P} \left( \|S_n^* - \tilde{S}_n\| > 3C(B_n m_n)^{1/3} \right) < \varepsilon.
\]
This proves the assertion of the lemma.

Step 2. It remains to show (A.11). For notational simplicity, we write $\delta_n$ and $\psi_n$ in place of $\delta_{C,n}$ and $\psi_{C,n}$, respectively. With $\sigma_n$ described in step 1, we consider the following functions on $\mathbb{R}^{m_n}$:
\[
g_n(x) \equiv \max \left\{ 0, 1 - d(x, A^{4\delta_n}) / \delta_n \right\}, \quad f_n(x) \equiv \mathbb{E} \left[ g_n(x + \sigma_n \mathcal{N}^*) \right],
\]
where $\mathcal{N}^*$ is an $m_n$-dimensional standard normal random vector and $d(x, A^{4\delta_n})$ denotes the distance between $x$ and the set $A^{4\delta_n}$. By Lemma 10.18 in Pollard (2001), $f_n(\cdot)$ is three-time continuously differentiable such that for all $(x, y)$,
\[
\left| f_n(x + y) - f_n(x) - \partial f_n(x)^\top y - \frac{1}{2} y^\top \partial^2 f_n(x) y \right| \leq \|y\| \frac{3}{\sigma_n^2 \delta_n},
\] (A.12)
and

$$(1 - \psi_n)1 \{x \in A\} \leq f_n(x) \leq \psi_n + (1 - \psi_n)1 \{x \in A^{3\delta_n}\}. \quad (A.13)$$

Let $\zeta_{n,t}, 1 \leq t \leq k_n + K_n$, be independent $m_n$-dimensional standard normal vectors and

$\tilde{\zeta}_{n,t} \equiv (V_{n,t}^*)^{1/2}\zeta_{n,t};$ recall the definition of $V_{n,t}^*$ from (A.1). We set

$$D_{n,t} \equiv \sum_{1 \leq s < t} Z_{n,s} + \sum_{t < s \leq k_n + K_n} \tilde{\zeta}_{n,s}.$$  

It is easy to see that

$$\int f_n(x)F_n(dx) = \mathbb{E}[f_n(D_{n,1} + \tilde{\zeta}_{n,1})], \quad \mathbb{E}[f_n(S_n^*)] = \mathbb{E}[f_n(D_{n,k_n + K_n} + Z_{n,k_n + K_n})],$$

and

$$D_{n,t} + Z_{n,t} = D_{n,t+1} + \tilde{\zeta}_{n,t+1}, \quad 1 \leq t \leq k_n + K_n - 1.$$  

Hence,

$$\mathbb{E}[f_n(S_n)] - \int f_n(x)F_n(dx) = \sum_{t=1}^{k_n + K_n} \left( \mathbb{E}[f_n(D_{n,t} + Z_{n,t})] - \mathbb{E}[f_n(D_{n,t} + \tilde{\zeta}_{n,t})] \right). \quad (A.14)$$

By (A.12), we have

$$\left| \mathbb{E}[f_n(D_{n,t} + Z_{n,t})] - \mathbb{E}[f_n(D_{n,t})] - \mathbb{E}[\partial f_n(D_{n,t})^\top Z_{n,t}] \right| - (1/2)\mathbb{E}[\mathbb{E}[(\partial^2 f_n(D_{n,t})Z_{n,t}Z_{n,t}^\top)|H_{n,t-1}]] \leq \frac{1}{\sigma_n^2\delta_n} \mathbb{E}[\|Z_{n,t}\|^3], \quad (A.15)$$

and

$$\left| \mathbb{E}[f_n(D_{n,t} + \tilde{\zeta}_{n,t})] - \mathbb{E}[f_n(D_{n,t})] - \mathbb{E}[\partial f_n(D_{n,t})^\top \tilde{\zeta}_{n,t}] \right| - (1/2)\mathbb{E}[\mathbb{E}[(\partial^2 f_n(D_{n,t})\tilde{\zeta}_{n,t}\tilde{\zeta}_{n,t}^\top)|H_{n,t-1}]] \leq \frac{1}{\sigma_n^2\delta_n} \mathbb{E}[\|\tilde{\zeta}_{n,t}\|^3]. \quad (A.16)$$

Since $\tilde{\zeta}_{n,t} = (V_{n,t}^*)^{1/2}\zeta_{n,t}$ and $\zeta_{n,t}$ is a standard normal random vector independent of $D_{n,t}$ and $V_{n,t}^*$, we have

$$\mathbb{E}[\partial f_n(D_{n,t})^\top \tilde{\zeta}_{n,t}] = 0 \quad \text{and} \quad \mathbb{E}[\text{Tr}(\partial^2 f_n(D_{n,t})\tilde{\zeta}_{n,t}\tilde{\zeta}_{n,t}^\top)] = \mathbb{E}[\text{Tr}(\partial^2 f_n(D_{n,t})V_{n,t}^*)]. \quad (A.17)$$

Let $\tilde{D}_{n,t} \equiv \sum_{1 \leq s < t} Z_{n,s} + (\Sigma_n - \sum_{s=1}^t V_{n,t}^*)\zeta_{n,t}$. We note that since $\Sigma_n$ is nonrandom, $\Sigma_n - \sum_{s=1}^t V_{n,t}^*$ is $\mathcal{H}_{n,t-1}$-measurable. We then observe that

$$\mathbb{E}[\partial f_n(D_{n,t})^\top Z_{n,t}] = \mathbb{E}[\partial f_n(\tilde{D}_{n,t})^\top Z_{n,t}] = \mathbb{E}[\partial f_n(\tilde{D}_{n,t})^\top \mathbb{E}[Z_{n,t}|\mathcal{H}_{n,t-1}, \zeta_{n,t}]] = \mathbb{E}[\partial f_n(\tilde{D}_{n,t})^\top \mathbb{E}[Z_{n,t}|\mathcal{H}_{n,t-1}]] = 0, \quad (A.18)$$

where the first equality holds because the conditional distribution of $\tilde{D}_{n,t}$ given $\mathcal{H}_{n,k_n + K_n}$ is the same as that of $D_{n,t}$; the second equality holds because $\sum_{1 \leq s < t} Z_{n,s}$ and $\Sigma_n - \sum_{s=1}^t V_{n,t}^*$ are
Similarly, \( \mathcal{H}_{n,t-1} \)-measurable; the third equality is by the independence between \( \zeta_{n,t} \) and \( \mathcal{H}_{n,K_n+K_n} \) and the last equality holds because \( (Z_{n,t}, \mathcal{H}_{n,t})_{1 \leq t \leq K_n+K_n} \) is a martingale difference array by construction. Similarly,

\[
E[\text{Tr}(\partial^2 f_n(D_{n,t})Z_{n,t}Z_{n,t}^\top)] = E[\text{Tr}(\partial^2 f_n(D_{n,t})Z_{n,t}Z_{n,t}^\top)] \\
= E[\text{Tr}(\partial^2 f_n(D_{n,t})E[Z_{n,t}Z_{n,t}^\top | \mathcal{H}_{n,t-1}, \zeta_{n,t}])] \\
= E[\text{Tr}(\partial^2 f_n(D_{n,t})E[Z_{n,t}Z_{n,t}^\top | \mathcal{H}_{n,t-1}])] \\
= E[\text{Tr}(\partial^2 f_n(D_{n,t})V_{n,t}^*]) = E[\text{Tr}(\partial^2 f_n(D_{n,t})V_{n,t}^*)]. \quad (A.19)
\]

Combining the results in (A.17), (A.18) and (A.19), we have

\[
E[\partial f_n(D_{n,t})^\top Z_{n,t}] = E[\partial f_n(D_{n,t})^\top \tilde{\zeta}_{n,t}] = 0 \\
E[\text{Tr}(\partial^2 f_n(D_{n,t})Z_{n,t}Z_{n,t}^\top)] = E[\text{Tr}(\partial^2 f_n(D_{n,t})\tilde{\zeta}_{n,t}\tilde{\zeta}_{n,t}^\top)].
\]

Combining this with (A.14), (A.15) and (A.16), we deduce

\[
\left| E[f(S_n^*)] - \int f_n(x) F_n(dx) \right| \\
\leq \frac{1}{\sigma_n^2 \delta_n} \sum_{t=1}^{K_n+K_n} \left( E[\|Z_{n,t}\|^3] + E[\|\tilde{\zeta}_{n,t}\|^3] \right) \\
= \frac{1}{\sigma_n^2 \delta_n} \sum_{t=1}^{K_n} \left( E[\|X_{n,t}\|^3] + E[\|V_{n,t}^{1/2} \zeta_{n,t} 1_{\{t \leq \tau_n\}}\|^3] \right) + \frac{2}{\sigma_n^2 \delta_n} \sum_{t=K_n+1}^{K_n+K_n} \mathbb{E}\left[\left\|K_n^{-1/2} \xi_{n,t}^{1/2} \eta_{n,t}\right\|^3\right] \\
\leq \frac{1}{\sigma_n^2 \delta_n} \sum_{t=1}^{K_n} \left( E[\|X_{n,t}\|^3] + E[\|V_{n,t}^{1/2} \zeta_{n,t}\|^3] \right) + \frac{2}{\sigma_n^2 \delta_n} \sum_{t=K_n+1}^{K_n+K_n} \mathbb{E}\left[\left\|K_n^{-1/2} \xi_{n,t}^{1/2} \eta_{n,t}\right\|^3\right] \\
\leq \frac{3B}{\sigma_n^2 \delta_n} + \frac{2}{\sigma_n^2 \delta_n K_n^{1/2}} \mathbb{E}\left[\left\|\xi_{n,t}^{1/2} N\right\|^3\right],
\]

where \( N \) is a generic \( m_n \)-dimensional standard normal random vector and the last inequality follows from (denoting by \( \Phi \) the distribution function of \( N \))

\[
E\left[\left\|Y_{n,t}^{1/2} \zeta_{n,t}\right\|^3\right] = E\left[\left(\int u^\top E[X_{n,t}X_{n,t}^\top | \mathcal{F}_{n,t-1}] u\right)^{3/2} \right] \\
= E\left[\int \left( u^\top E[X_{n,t}X_{n,t}^\top | \mathcal{F}_{n,t-1}] u\right)^{3/2} \Phi(du) \right] \\
= E\left[\int \left( E[(u^\top X_{n,t})^2 | \mathcal{F}_{n,t-1}] \right)^{3/2} \Phi(du) \right] \\
\leq E\left[\int E[\left\|u^\top X_{n,t}\right\|^3 | \mathcal{F}_{n,t-1}] \Phi(du) \right] \\
= E\left[\left\|\zeta_{n,t}^\top X_{n,t}\right\|^3 \right] = \sqrt{8/\pi} \mathbb{E}\left[\left\|X_{n,t}\right\|^3\right].
\]
Note that $\Sigma_n - \xi_n$ is positive semi-definite. Hence,

$$
\mathbb{E}\left[\left\|\xi_n^{1/2} N\right\|^3\right] = \mathbb{E}\left[(N^\top \xi_n N)^{3/2}\right] \leq \mathbb{E}\left[(N^\top \Sigma_n N)^{3/2}\right] \\
\leq \lambda_{\text{max}}(\Sigma_n)^{3/2} \mathbb{E}\left[(N^\top N)^{3/2}\right] \\
\leq \tilde{\lambda}^{3/2} \left(\mathbb{E}\left[(N^\top N)^2\right]\right)^{3/4} \leq 3\tilde{\lambda}^{3/2} m_n^{3/2}.
$$

Hence, under the condition $K_n \geq 36\tilde{\lambda}^3 m_n^3 / B_n^2$,

$$
\left|\mathbb{E}[f(S_n^*)] - \int f_n(x) F_n(dx)\right| \leq \frac{3B_n}{\sigma_n^2 \delta_n} + \frac{6\tilde{\lambda}^{3/2} m_n^{3/2}}{\sigma_n^2 \delta_n K_n^{1/2}} \leq \frac{4B_n}{\sigma_n^2 \delta_n}. \quad (A.20)
$$

From (A.13) and (A.20),

$$
\mathbb{P}(S_n^* \in A) \leq \frac{1}{1 - \psi_n} \mathbb{E}[f_n(S_n^*)] \\
\leq \frac{1}{1 - \psi_n} \left(\int f_n(x) F_n(dx) + \frac{4B_n}{\sigma_n^2 \delta_n}\right) \\
\leq \frac{1}{1 - \psi_n} \left(\psi_n + (1 - \psi_n) F_n(A^{3\delta_n}) + \frac{4B_n}{\sigma_n^2 \delta_n}\right) \\
= F_n(A^{3\delta_n}) + \frac{1}{1 - \psi_n} \left(\psi_n + \frac{4B_n}{\sigma_n^2 \delta_n}\right),
$$

which proves (A.11) as wanted. \hfill Q.E.D.

**Proof of Theorem 1.** Let $K_n$ satisfy the condition in Lemma A2 and then define $S_n^*$ as in (A.2). The assertion of Theorem 1 then readily follows from Lemma A1 and Lemma A2. \hfill Q.E.D.

**S.A.2 Proof of Theorem 2**

**Proof of Theorem 2.** Step 1. The proof for part (a) of the theorem is divided into 3 steps. Below, for a generic real sequence $a_n$, let $O_{pu}(a_n)$ denote a random sequence that is $O_p(a_n)$ uniformly in $x \in \mathcal{X}$. In this step, we show

$$
\frac{n^{1/2}P(x)^\top (\hat{b}_n - b_n^*)}{\sigma_n(x)} = \frac{n^{-1/2}P(x)^\top Q_n^{-1} P_n^\top U_n}{\sigma_n(x)} + O_{pu}(\delta_{1,n} + m_n^{1/2}\delta_{3,n}), \quad (A.21)
$$

where $P_n \equiv [P(X_1), \ldots, P(X_n)]^\top$ and $U_n = (u_1, \ldots, u_n)^\top$.

By Assumption 2(ii),

$$
\sup_{x \in \mathcal{X}} \frac{\|P(x)\|}{\sigma_n(x)} \leq (\lambda_{\text{min}}(A_n))^{-1/2} \lambda_{\text{max}}(Q_n) \leq K. \quad (A.22)
$$
By Assumptions 2(ii), (iv) and (v), we have, with probability approaching one,

\[ \lambda_{\text{max}}(\hat{Q}_n) + \lambda_{\text{max}}(\hat{A}_n) + \lambda_{\text{min}}^{-1}(\hat{Q}_n) + \lambda_{\text{min}}^{-1}(\hat{A}_n) \leq K. \]  

(A.23)

Let \( h_n^* (\cdot) \equiv P (\cdot)^T b_n^* \), \( H_n = (h(X_1), \ldots, h(X_n))^T \) and \( H_n^* = (h_n^*(X_1), \ldots, h_n^*(X_n))^T \). By the definition of \( b_n \), we can decompose

\[ \hat{b}_n - b_n^* = (P_n^T P_n)^{-1} \left( P_n^T U_n \right) + (P_n^T P_n)^{-1} P_n^T (H_n - H_n^*). \]  

(A.24)

By Assumption 2(ii),

\[ \mathbb{E} \left[ \left\| n^{-1/2} P_n^T U_n \right\|^{2} \right] = n^{-1} \text{Tr} (A_n) \leq K m_n n^{-1}, \]  

(A.25)

which together with Markov’s inequality implies that

\[ \left\| n^{-1/2} P_n^T U_n \right\| = O_p(m_n^{1/2}). \]  

(A.26)

We observe

\[ \sup_{x \in \mathcal{X}} \frac{1}{\sigma_n (x)} \left| n^{1/2} P(x)^T (P_n^T P_n)^{-1} \left( P_n^T U_n \right) - n^{-1/2} P(x)^T Q_n^{-1} \left( P_n^T U_n \right) \right| \]

\[ = \sup_{x \in \mathcal{X}} \frac{1}{\sigma_n (x)} \left| P(x)^T \hat{Q}_n^{-1} \left( \hat{Q}_n - Q_n \right) Q_n^{-1} \left( n^{-1/2} P_n^T U_n \right) \right| \]

\[ \leq (\lambda_{\text{min}}(\hat{Q}_n) \lambda_{\text{min}}(Q_n))^{-1} \left\| \hat{Q}_n - Q_n \right\| \sup_{x \in \mathcal{X}} \frac{\|P(x)\|}{\sigma_n (x)} \]

\[ = O_p(m_n^{1/2} \delta_{3,n}), \]

where the inequality follows from the Cauchy–Schwarz inequality and the last line is derived from Assumption 2(iv), (A.22), (A.23) and (A.26).

By Assumption 2(ii), (A.22) and (A.23),

\[ \sup_{x \in \mathcal{X}} \frac{1}{\sigma_n (x)} \left| n^{1/2} P(x)^T (P_n^T P_n)^{-1} P_n^T (H_n - H_n^*) \right| \]

\[ \leq \sup_{x \in \mathcal{X}} \frac{\|P(x)\|}{\sigma_n (x)} \left( (H_n - H_n^*)^T P_n (P_n^T P_n)^{-1/2} \hat{Q}_n^{-1/2} (P_n^T P_n)^{-1/2} P_n^T (H_n - H_n^*) \right)^{1/2} \]

\[ \leq \left( (H_n - H_n^*)^T P_n (P_n^T P_n)^{-1} P_n^T (H_n - H_n^*) \right)^{1/2} \sup_{x \in \mathcal{X}} \frac{\|P(x)\|}{\sigma_n (x)} \]

\[ \leq \sup_{x \in \mathcal{X}} n^{1/2} \left| h(x) - P(x)^T b_n^* \right| \frac{\|P(x)\|}{\sigma_n (x)} = O_p(\delta_{1,n}). \]  

(A.28)

The claim in (A.21) follows by combining the results in (A.24), (A.27) and (A.28).

**Step 2.** In this step, we show that

\[ \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} P(x)^T (\hat{b}_n - b_n^*)}{\hat{\sigma}_n (x)} - \frac{n^{1/2} P(x)^T (\hat{b}_n - b_n^*)}{\hat{\sigma}_n (x)} \right| = O_p(m_n^{1/2}(\delta_{3,n} + \delta_{4,n})). \]  

(A.29)
By the triangle inequality
\[ \|\widehat{\Sigma}_n - \Sigma_n\|_S \leq \|(\widehat{Q}_n^{-1} - Q_n^{-1})\widehat{A}_n\widehat{Q}_n^{-1}\|_S + \|Q_n^{-1}(\widehat{A}_n - A_n)\widehat{Q}_n^{-1}\|_S + \|Q_n^{-1}A_n(\widehat{Q}_n^{-1} - Q_n^{-1})\|_S. \]

By the Cauchy–Schwarz inequality, Assumption 2(ii, iv) and (A.23),
\[ \|(\widehat{Q}_n^{-1} - Q_n^{-1})\widehat{A}_n\widehat{Q}_n^{-1}\|_S \leq \frac{\lambda_{\max}(\widehat{A}_n)\|\widehat{Q}_n - Q_n\|_S}{\lambda_{\min}(Q_n)(\lambda_{\min}(Q_n))^2} = O_p(\delta_{3,n}). \]

Similarly, \(\|Q_n^{-1}(\widehat{A}_n - A_n)\widehat{Q}_n^{-1}\|_S = O_p(\delta_{4,n})\) and \(\|Q_n^{-1}A_n(\widehat{Q}_n^{-1} - Q_n^{-1})\|_S = O_p(\delta_{3,n})\). Combining these estimates, we get
\[ \|\widehat{\Sigma}_n - \Sigma_n\|_S = O_p(\delta_{3,n} + \delta_{4,n}). \tag{A.30} \]

By Assumption 2(ii), this estimate further implies that, with probability approaching one,
\[ \lambda_{\min}^{-1}(\widehat{\Sigma}_n) \leq K, \quad \lambda_{\max}(\widehat{\Sigma}_n) \leq K. \tag{A.31} \]

We then observe
\[
\sup_{x \in \mathcal{X}} \frac{|\sigma_n(x) - \hat{\sigma}_n(x)|}{\hat{\sigma}_n(x)} = \sup_{x \in \mathcal{X}} \frac{|\sigma_n^2(x) - \hat{\sigma}_n^2(x)|}{\hat{\sigma}_n(x)(\sigma_n(x) + \hat{\sigma}_n(x))} \\
= \sup_{x \in \mathcal{X}} \frac{|P(x)\top(\widehat{\Sigma}_n - \Sigma_n)P(x)|}{\hat{\sigma}_n(x)(\sigma_n(x) + \hat{\sigma}_n(x))} \\
\leq \frac{\|\widehat{\Sigma}_n - \Sigma_n\|_S}{(\lambda_{\min}(\widehat{\Sigma}_n)\lambda_{\min}(\Sigma_n))^{1/2} + \lambda_{\min}(\Sigma_n)} = O_p(\delta_{3,n} + \delta_{4,n}) \tag{A.32}
\]

where the last line follows from Assumption 2(ii), (A.30) and (A.31).

By the Cauchy–Schwarz inequality
\[ \sup_{x \in \mathcal{X}} \left| \frac{P(x)}{\sigma_n(x)} 1^n P(x) \right| \leq \left\| \frac{1^n P(x)}{\lambda_{\min}(Q_n)} \right\| \sup_{x \in \mathcal{X}} \|P(x)\| = O_p(m_n^{1/2}) \tag{A.33} \]

where the equality is due to Assumption 2(ii), (A.22) and (A.26). By (A.21) and (A.33),
\[ \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2}P(x)\top(\hat{b}_n - b_n^*)}{\sigma_n(x)} \right| = O_p(m_n^{1/2}). \tag{A.34} \]

Combining (A.32) and (A.34), we deduce
\[ \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2}P(x)\top(\hat{b}_n - b_n^*)}{\sigma_n(x)} - \frac{n^{1/2}P(x)\top(\hat{b}_n - b_n^*)}{\sigma_n(x)} \right| \leq \sup_{x \in \mathcal{X}} \frac{|\sigma_n(x) - \hat{\sigma}_n(x)|}{\hat{\sigma}_n(x)} \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2}P(x)\top(\hat{b}_n - b_n^*)}{\sigma_n(x)} \right| = O_p(m_n^{1/2}(\delta_{3,n} + \delta_{4,n})), \]
which finishes the proof of (A.29).

Step 3. In this step, we show the assertion in part (a) of the theorem. It suffices to show that,

\[
\frac{n^{1/2} \left( \hat{b}_n(x) - h(x) \right)}{\hat{\sigma}_n(x)} = \frac{P(x)^\top \tilde{S}_n}{\sigma_n(x)} + O_p(\delta_n), \tag{A.35}
\]

where \( \tilde{S}_n \equiv n^{-1} \tilde{N}_n \) is \( \mathcal{N}(0, \Sigma_n) \) distributed.

By (A.21) and (A.29),

\[
\frac{n^{1/2}P(x)^\top (\hat{b}_n - b_n^*)}{\hat{\sigma}_n(x)} = \frac{n^{-1/2}P(x)^\top Q_n^{-1}P_n^\top U_n}{\sigma_n(x)} + O_p(\delta_1,n + m_n^{1/2}(\delta_3,n + \delta_4,n)). \tag{A.36}
\]

By Assumption 2(ii) and (A.22),

\[
\sup_{x \in \mathcal{X}} \left\| \frac{P(x)^\top Q_n^{-1}}{\sigma_n(x)} \right\| \leq K. \tag{A.37}
\]

Hence, by Assumption 2(iii), (A.36) and (A.37),

\[
\frac{n^{1/2}P(x)^\top (\hat{b}_n - b_n^*)}{\hat{\sigma}_n(x)} = \frac{P(x)^\top Q_n^{-1} \tilde{N}_n}{\sigma_n(x)} + O_p(\delta_1,n + \delta_2,n + m_n^{1/2}(\delta_3,n + \delta_4,n)). \tag{A.38}
\]

By Assumption 2(i) and (A.22),

\[
\sup_{x \in \mathcal{X}} \left| \frac{n^{1/2}(h(x) - P(x)^\top b_n^*)}{\sigma_n(x)} \right| = \sup_{x \in \mathcal{X}} \left\| P(x) \right\| \left| \frac{n^{1/2}(h(x) - P(x)^\top b_n^*)}{\sigma_n(x)} \right| = O(\delta_1,n), \tag{A.39}
\]

which combined with (A.32) yields

\[
\sup_{x \in \mathcal{X}} \left| \frac{n^{1/2}(h(x) - P(x)^\top b_n^*)}{\hat{\sigma}_n(x)} \right| \leq \sup_{x \in \mathcal{X}} \left| \frac{\hat{\sigma}_n(x) - \sigma_n(x)}{\hat{\sigma}_n(x)} \right| \left| \frac{n^{1/2}(h(x) - P(x)^\top b_n^*)}{\sigma_n(x)} \right| + \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2}(h(x) - P(x)^\top b_n^*)}{\sigma_n(x)} \right| = O_p(\delta_1,n). \tag{A.40}
\]

From (A.38) and (A.40), the assertion (A.35) readily follows.

Step 4. Given the result in part (a), the assertion of part (b) can be shown by using similar arguments in the proof of Theorem 5.6 in Belloni, Chernozhukov, Chetverikov, and Kato (2015). We omit the proof for brevity.

Q.E.D.

S.A.3 Proof of Theorem 3

We first introduce some notations and a preliminary estimate; see Lemma A3, below. Recall that the feasible estimator \( \hat{b}_n \) is given by

\[
\hat{b}_n \equiv \left( \sum_{t=1}^{n} P(X_t) P(X_t)^\top \right)^{-1} \left( \sum_{t=1}^{n} P(X_t) g(Y_t^* ; \hat{\gamma}_n) \right).
\]

11
We denote the corresponding infeasible estimator by

\[ \hat{b}_n^\dagger = \left( \sum_{t=1}^{n} P(X_t) P(X_t)^\top \right)^{-1} \left( \sum_{t=1}^{n} P(X_t) g(Y_t^*, \gamma_0) \right). \]

**Lemma A3.** Let \( \delta_{6,n} \equiv \sup_{x \in \mathcal{X}} \| P(x) \|^{-1} \). Under Assumption 4,

\[ \sup_{x \in \mathcal{X}} \frac{n^{1/2} P(x)^\top (\hat{b}_n - \hat{b}_n^\dagger)}{\sigma_n(x)} = O_p(\delta_{3,n} + \delta_{5,n} + \delta_{6,n}). \] (A.41)

**Proof of Lemma A3.** Step 1. In this step, we show that

\[ \sup_{x \in \mathcal{X}} \frac{n^{1/2} P(x)^\top (\hat{b}_n - \hat{b}_n^\dagger)}{\sigma_n(x)} = \sup_{x \in \mathcal{X}} \frac{n^{1/2} P(x)^\top Q_n^{-1} G_n (\hat{\gamma}_n - \gamma_0)}{\sigma_n(x)} + O_p(\delta_{3,n} + \delta_{5,n}). \]

By definition,

\[ \hat{b}_n - \hat{b}_n^\dagger = \left( P_n^\top P_n \right)^{-1} \sum_{t=1}^{n} P(X_t) (g(Y_t^*, \hat{\gamma}_n) - g(Y_t^*, \gamma_0)). \] (A.42)

Applying the second-order Taylor expansion, we further deduce

\[ \hat{b}_n - \hat{b}_n^\dagger = \left( P_n^\top P_n \right)^{-1} \sum_{t=1}^{n} P(X_t) g_\gamma(Y_t^*, \gamma_0)^\top (\hat{\gamma}_n - \gamma_0) \]

\[ + \frac{1}{2} \left( P_n^\top P_n \right)^{-1} \sum_{t=1}^{n} P(X_t) (\hat{\gamma}_n - \gamma_0)^\top g_{\gamma\gamma}(Y_t^*, \hat{\gamma}_n) (\hat{\gamma}_n - \gamma_0), \] (A.43)

where \( \hat{\gamma}_n \) is a mean value between \( \hat{\gamma}_n \) and \( \gamma_0 \) that may vary across rows. By Assumption 4(iv,vi),

\[ n^{-1} \sum_{t=1}^{n} (\hat{\gamma}_n - \gamma_0)^\top g_{\gamma\gamma}(Y_t^*, \hat{\gamma}_n) (\hat{\gamma}_n - \gamma_0)^2 \]

\[ = O_p(n^{-2}). \] (A.44)

Since \( P_n \left( P_n^\top P_n \right)^{-1} P_n^\top \) is a projection matrix,

\[ \left\| \left( P_n^\top P_n \right)^{-1} \sum_{t=1}^{n} P(X_t) (\hat{\gamma}_n - \gamma_0)^\top g_{\gamma\gamma}(Y_t^*, \hat{\gamma}_n) (\hat{\gamma}_n - \gamma_0) \right\|^2 \]

\[ \leq \lambda_{\text{min}}^{-1}(\hat{Q}_n) n^{-1} \sum_{t=1}^{n} (\hat{\gamma}_n - \gamma_0)^\top g_{\gamma\gamma}(Y_t^*, \hat{\gamma}_n) (\hat{\gamma}_n - \gamma_0)^2 \]

\[ = O_p(n^{-2}), \] (A.45)

where the rate statement follows from (A.44) and Assumption 3. Collecting the results in (A.43) and (A.45), we get

\[ \hat{b}_n - \hat{b}_n^\dagger = \left( P_n^\top P_n \right)^{-1} \sum_{t=1}^{n} P(X_t) g_\gamma(Y_t^*, \gamma_0)^\top (\hat{\gamma}_n - \gamma_0) + O_p(n^{-1}). \] (A.46)
By Assumptions 3 and 4(ii,vi),
\begin{equation}
\left(P_n^\top P_n\right)^{-1} \sum_{t=1}^{n} P(X_t)g_\gamma(Y_t^*, \gamma_0)^\top (\hat{\gamma}_n - \gamma_0) = \hat{Q}_n^{-1}G_n(\hat{\gamma}_n - \gamma_0) + O_p(\delta_{3,n}n^{-1/2}). \tag{A.47}
\end{equation}
For $1 \leq j \leq d$, let $g_{\gamma,j}(Y_t^*, \gamma_0)$ denote the $j$th component of $g_\gamma(Y_t^*, \gamma_0)$ and let $G_{n,j}$ denote the $j$th column of $G_n$. We note that
\begin{equation}
G_{n,j}^\top Q_n^{-1}G_{n,j} \leq n^{-1} \sum_{t=1}^{n} \mathbb{E} \left[ g_{\gamma,j}(Y_t^*, \gamma_0)^2 \right], \tag{A.48}
\end{equation}
because the left-hand side is the squared $L_2$-norm of the projection of $g_{\gamma,j}(Y_t^*, \gamma_0)$ onto the column space of $P(X_t)$ under the product measure $\mathbb{P} \otimes P_n$, with $P_n$ being the empirical measure. Hence, for any $1 \leq j \leq d$,
\begin{equation}
\|G_{n,j}\|^2 \leq \lambda_{\max}(Q_n) \sum_{t=1}^{n} \mathbb{E} \left[ g_{\gamma,j}(Y_t^*, \gamma_0)^2 \right] \leq \lambda_{\max}(Q_n) \sup_t \mathbb{E} \left[ \|g_\gamma(Y_t^*, \gamma_0)\|^2 \right]. \tag{A.49}
\end{equation}
By Assumptions 2(ii) and 4(iv), we further deduce
\begin{equation}
\|G_n\|^2 \leq K. \tag{A.50}
\end{equation}
By Assumption 3, Assumption 4(iii) and (A.50),
\begin{align*}
\left\| \left( \hat{Q}_n^{-1} - Q_n^{-1} \right) G_n(\hat{\gamma}_n - \gamma_0) \right\|^2 &= \left\| \hat{Q}_n^{-1} \left( \hat{Q}_n^{-1} - Q_n^{-1} \right) Q_n^{-1} G_n(\hat{\gamma}_n - \gamma_0) \right\|^2 \\
&\leq \frac{\|\hat{\gamma}_n - \gamma_0\|^2}{(\lambda_{\min}(Q_n)\lambda_{\min}(Q_n))^2} \left\| \hat{Q}_n^{-1} - Q_n^{-1} \right\|^2 \|G_n\|^2 = O_p(n^{-1} \delta_{3,n}^2),
\end{align*}
which further implies that
\begin{equation}
\left( \hat{Q}_n^{-1} - Q_n^{-1} \right) G_n(\hat{\gamma}_n - \gamma_0) = O_p(n^{-1/2} \delta_{3,n}). \tag{A.51}
\end{equation}
Combining the results in (A.47) and (A.51), we get
\begin{equation}
\hat{b}_n - \hat{b}_n^\dagger = Q_n^{-1}G_n(\hat{\gamma}_n - \gamma_0) + O_p(\delta_{3,n} + \delta_{5,n})n^{-1/2}). \tag{A.52}
\end{equation}
With an appeal to the Cauchy–Schwarz inequality, we deduce (A.41) from (A.52) and (A.22).

**Step 2.** We now prove the assertion of Lemma A3. For $j \in \{1, \ldots, d\}$, let $\phi_{n,j}^*$ denote the $j$th column of $\phi_n^*$; recall the definition of $\phi_n^*$ from Assumption 4. By definition,
\begin{align}
P(x)^\top Q_n^{-1}G_{n,j} - P(x)^\top \phi_{n,j}^* &\nonumber \\
&= P(x)^\top Q_n^{-1} (G_{n,j} - Q_n\phi_{n,j}) \\
&= n^{-1} \sum_{t=1}^{n} P(x)^\top Q_n^{-1} \left( \mathbb{E}[P(X_t)(H_j(X_t) - H_j^*(X_t))] \right), \tag{A.53}
\end{align}
where \( H_j(X_t) = \mathbb{E}[g_{\gamma,j}(Y_t^*, \gamma_0) | X_t] \) and \( H_j^*(X_t) = P(X_t)^T \phi_n^* \) (we suppress the dependence of these functions on \( n \) for notational simplicity). Using similar arguments that lead to \( (A.48) \), we can show that

\[
\left\| \mathbb{E} \left[ n^{-1} \sum_{t=1}^n P(X_t) (H_j(X_t) - H_j^*(X_t)) \right] \right\|^2 \leq \lambda_{\text{max}}(Q_n) \mathbb{E} \left[ \left\| n^{-1} \sum_{t=1}^n (H_j(X_t) - H_j^*(X_t)) \right\|^2 \right],
\]

which together with Assumption 3 and Assumption 4(iii) implies that

\[
\left\| \mathbb{E} \left[ n^{-1} \sum_{t=1}^n P(X_t) (H_j(X_t) - H_j^*(X_t)) \right] \right\| = O(m_n^{-\rho}). \tag{A.54}
\]

By \( (A.54) \) and the Cauchy–Schwarz inequality,

\[
\sup_{x \in \mathcal{X}} \left\| n^{-1} \sum_{t=1}^n P(x)^T Q_n^{-1} \mathbb{E} \left[ P(X_t) (H_j(X_t) - H_j^*(X_t)) \right] \right\|^2 \leq \sup_{x \in \mathcal{X}} \| P(x) \|^2 (\lambda_{\min}(Q_n))^{-2} \mathbb{E} \left[ \left\| n^{-1} \sum_{t=1}^n P(X_t) (H_j(X_t) - H_j^*(X_t)) \right\|^2 \right] \leq Km_n^{1-2\rho} \sigma_n^2 = O(1)
\]

where the last line is by \( (A.54) \), Assumption 3 and Assumption 4(v). By \( (A.53) \), we further deduce that

\[
\sup_{x \in \mathcal{X}} \left\| P(x)^T Q_n^{-1} G_n - P(x)^T \phi_n^* \right\| \leq K. \tag{A.55}
\]

From Assumption 4(iii,iv), it is easy to see that \( P(\cdot)^T \phi_n^* \) is uniformly bounded. Hence, by \( (A.55) \),

\[
\sup_{x \in \mathcal{X}} \left\| P(x)^T Q_n^{-1} G_n \right\| \leq K. \tag{A.56}
\]

Using the Cauchy–Schwarz inequality, we deduce from \( (A.56) \), Assumption 3 and Assumption 4(vi) that

\[
\sup_{x \in \mathcal{X}} \left\| n^{1/2} P(x)^T Q_n^{-1} G_n (\hat{\gamma}_n - \gamma_0) \right\| \left\| \frac{\sigma_n(x)}{\sigma_n(x)} \right\| \leq \lambda_{\max}(Q_n)(\lambda_{\min}(A_n))^{-1/2} \left\| n^{1/2}(\hat{\gamma}_n - \gamma_0) \right\| \sup_{x \in \mathcal{X}} \frac{\| P(x)^T Q_n^{-1} G_n \|}{\| P(x) \|} = O_p(\delta_{6,n}).
\]

The assertion of Lemma A3 then follows from this estimate and \( (A.41) \). \( \text{Q.E.D.} \)

**Proof of Theorem 3.** By Assumption 3, we can invoke \( (A.32) \) in the proof of Theorem 2 to get

\[
\sup_{x \in \mathcal{X}} \left| \frac{\sigma_n(x) - \hat{\sigma}_n(x)}{\hat{\sigma}_n(x)} \right| = O_p(\delta_{4,n} + \delta_{1,n}), \tag{A.57}
\]
which together with Lemma A3 implies that

\[
\sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} P(x)\hat{\sigma}_n (x) - n^{1/2} P(x)\tilde{\sigma}_n (x)}{\sigma_n (x)} \right| \\
\leq \sup_{x \in \mathcal{X}} \frac{\sigma_n (x) - \tilde{\sigma}_n (x)}{\sigma_n (x)} \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} P(x)\tilde{\sigma}_n (x)}{\sigma_n (x)} \right| \\
= O_p((\delta_{3,n} + \delta_{5,n} + \delta_{6,n})(\delta_{3,n} + \delta_{4,n})).
\]

Therefore,

\[
\sup_{x \in \mathcal{X}} \frac{n^{1/2} P(x)\tilde{\sigma}_n (x)}{\sigma_n (x)} = O_p(\delta_{3,n} + \delta_{5,n} + \delta_{6,n}).
\]

Let \( \hat{\sigma}_n (x) = P(x)\sigma_n (x) \). Applying (A.35) with \( \hat{\sigma}_n (x) \) replacing \( \tilde{\sigma}_n (x) \),

\[
\frac{n^{1/2} (\hat{\sigma}_n (x) - \sigma_n (x))}{\sigma_n (x)} = \frac{P(x)\hat{\sigma}_n (x)}{\sigma_n (x)} + O_p(\delta_n).
\]

Then, by Lemma A3,

\[
\frac{n^{1/2} (\hat{\sigma}_n (x) - \sigma_n (x))}{\sigma_n (x)} = \frac{P(x)\hat{\sigma}_n (x)}{\sigma_n (x)} + O_p(\delta_n).
\]

Under the null hypothesis (2.9) with \( h(x) = 0 \),

\[
\sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} \hat{\sigma}_n (x)}{\sigma_n (x)} \right| = \sup_{x \in \mathcal{X}} \left| \frac{P(x)\hat{\sigma}_n (x)}{\sigma_n (x)} \right| + O_p(\delta_n + \delta_{5,n} + \delta_{6,n}).
\]

Note that \( (\delta_n + \delta_{5,n} + \delta_{6,n}) \log m_n)^{1/2} = o(1) \) under maintained assumptions. The first assertion in Theorem 3 then follows from (A.61) and the argument in the proof of Theorem 5.6 in Belloni, Chernozhukov, Chetverikov, and Kato (2015).

We now turn to the second assertion. By the triangle inequality, (A.57) and (A.60),

\[
\sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} \hat{\sigma}_n (x)}{\sigma_n (x)} \right| \geq \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} h(x)}{\sigma_n (x)} \right| - \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} (\hat{\sigma}_n (x) - \sigma_n (x))}{\sigma_n (x)} \right| \\
= \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} h(x)}{\sigma_n (x)} \sigma_n (x) \right| - \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} (\hat{\sigma}_n (x) - \sigma_n (x))}{\sigma_n (x)} \right| \\
\geq \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} h(x)}{\sigma_n (x)} \right| \left( 1 - \sup_{x \in \mathcal{X}} \left| \frac{\sigma_n (x) - 1}{\tilde{\sigma}_n (x)} \right| \right) - \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} (\hat{\sigma}_n (x) - \sigma_n (x))}{\sigma_n (x)} \right| \\
= \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} h(x)}{\sigma_n (x)} \right| (1 + o_p(1)) - \sup_{x \in \mathcal{X}} \left| \frac{P(x)\tilde{\sigma}_n (x)}{\sigma_n (x)} \right| + o_p(1).
\]
By the Cauchy–Schwarz inequality and Assumption 3,
\[
\sup_{x \in \mathcal{X}} \left| \frac{P(x) \hat{S}_n}{\sigma_n(x)} \right| \leq \frac{\lambda_{\max}(Q_n)}{(\lambda_{\min}(A_n))^{1/2}} \left\| \hat{S}_n \right\| = O_p(m_n^{1/2}). \tag{A.63}
\]
Since \( \mathbb{E}[g(Y_t^*, \gamma_0)|X_t = x] \neq 0 \) for some \( x \in \mathcal{X} \), there exists some constant \( c_0 > 0 \) such that \( \sup_{x \in \mathcal{X}} |h(x)| > c_0 \). Moreover, by Assumption 4(v), \( \sup_{x \in \mathcal{X}} \|P(x)\| \leq \zeta_n m_n^{1/2} \). Therefore,
\[
\sup_{x \in \mathcal{X}} \left| n^{1/2}h(x) \right| / \sigma_n(x) \geq \frac{n^{1/2} \lambda_{\min}(Q_n) c_0}{(\lambda_{\max}(A_n))^{1/2} \zeta_n m_n^{1/2}}, \tag{A.64}
\]
which together with (A.62) and (A.63) implies that (recalling \( \zeta_n m_n \ll n^{1/2} \) from Assumption 4(vii)),
\[
\sup_{x \in \mathcal{X}} \left| n^{1/2} \hat{h}(x) \right| / \sigma_n(x) \geq \frac{n^{1/2} \lambda_{\min}(Q_n) c_0}{(\lambda_{\max}(A_n))^{1/2} \zeta_n m_n^{1/2}} (1 + o_p(1)). \tag{A.65}
\]
Like (A.73) in Belloni, Chernozhukov, Chetverikov, and Kato (2015), we can show that the critical value \( cv_{n, \alpha} \) satisfies \( cv_{n, \alpha} = O_p((\log(m_n))^{1/2}) \), which together with Assumptions 3(ii) and 4(vii) implies that
\[
\frac{\lambda_{\max}(A_n)^{1/2} cv_{n, \alpha} \zeta_n m_n^{1/2}}{\lambda_{\min}(Q_n)(1 + o_p(1)) m_n^{1/2}} = o_p(1). \tag{A.66}
\]
Combining the results in (A.65) and (A.66), we deduce that
\[
\Pr \left( \hat{T}_n \leq cv_{n, \alpha} \right) \leq \Pr \left( \frac{n^{1/2} \lambda_{\min}(Q_n) c_0 (1 + o_p(1))}{(\lambda_{\max}(A_n))^{1/2} \zeta_n m_n^{1/2}} \leq cv_{n, \alpha} \right) = \Pr \left( c_0 \leq \frac{\lambda_{\max}(A_n)^{1/2} cv_{n, \alpha} \zeta_n m_n^{1/2}}{\lambda_{\min}(Q_n)(1 + o_p(1)) m_n^{1/2}} \right) \to 0
\]
as \( n \to \infty \). From here, the second assertion readily follows.

**Q.E.D.**

### S.A.4 Proof of Theorem 4

We first establish the martingale approximation as claimed in (3.2) and (3.3) in the main text; see Lemma A4 below. The variables \( X_{n,t}^* \) and \( \tilde{X}_{n,t} \) are defined as follows:
\[
X_{n,t}^* = \sum_{s=-\infty}^{\infty} \{ \mathbb{E}[X_{n,t+s}|\mathcal{F}_{n,t}] - \mathbb{E}[X_{n,t+s}|\mathcal{F}_{n,t-1}] \}, \tag{A.67}
\]
\[
\tilde{X}_{n,t} \equiv \sum_{s=0}^{\infty} \mathbb{E}[X_{n,t+s}|\mathcal{F}_{n,t-1}] - \sum_{s=0}^{\infty} \{X_{t-s-1} - \mathbb{E}[X_{t-s-1}|\mathcal{F}_{n,t-1}] \} \tag{A.68}
\]

**Lemma A4.** The following statements hold under Assumption 5 for each \( j \in \{1, \ldots, m_n \} \)
\[
(a) \sum_{s=-\infty}^{\infty} \left\| \mathbb{E} \left[ X_{n,t+s}^{(j)}|\mathcal{F}_{n,t} \right] - \mathbb{E} \left[ X_{n,t+s}^{(j)}|\mathcal{F}_{n,t-1} \right] \right\|_q \leq 4 \bar{c}_n \kappa_n^{-1/2} \sum_{s=0}^{\infty} \psi_s; 
\]

16
(b) \( \sup_{j,t,n} \|X^{s(j)}_{n,t}\| \leq 4\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s \) and \( \mathbb{E} \left[ X^*_n | \mathcal{F}_{n,t-1} \right] = 0; \)
(c) \( \sum_{s=0}^{\infty} \left\| \mathbb{E} \left[ X^{s(j)}_{n,t+s} | \mathcal{F}_{n,t-1} \right] \right\|_q + \sum_{s=0}^{\infty} \left\| X^{s(j)}_{t-s-1} - \mathbb{E} \left[ X^{s(j)}_{t-s-1} | \mathcal{F}_{n,t-1} \right] \right\|_q \leq 2\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s; \)
(d) \( \sup_{j,t,n} \|X^{s(j)}_{n,t}\| < 2\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s; \)
(e) \( X_{n,t} = X^*_n + \bar{X}_{n,t} - \bar{X}_{n,t+1} \) for each \( t \geq 1; \)
(f) \( \| S_n - S^*_n \|_2 = O(\bar{c}_n m_n^{-1/2} k_n^{-1/2}). \)

**Proof of Lemma A4.** (a) We first note that

\[
\sum_{s=0}^{\infty} \left\| \mathbb{E} \left[ X^{s(j)}_{n,t+s} | \mathcal{F}_{n,t} \right] - \mathbb{E} \left[ X^{s(j)}_{n,t+s} | \mathcal{F}_{n,t-1} \right] \right\|_q \\
\leq \sum_{s=0}^{\infty} \left\| \mathbb{E} \left[ X^{s(j)}_{n,t+s} | \mathcal{F}_{n,t} \right] \right\|_q + \sum_{s=0}^{\infty} \left\| \mathbb{E} \left[ X^{s(j)}_{n,t+s} | \mathcal{F}_{n,t-1} \right] \right\|_q \\
\leq \bar{c}_n k_n^{-1/2} \left( \sum_{s=0}^{\infty} \psi_s + \sum_{s=0}^{\infty} \psi_{s+1} \right) \\
\leq 2\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s < \infty.
\]

In addition, we have

\[
\sum_{s=-\infty}^{-1} \left\| \mathbb{E} \left[ X^{s(j)}_{n,t+s} | \mathcal{F}_{n,t} \right] - \mathbb{E} \left[ X^{s(j)}_{n,t+s} | \mathcal{F}_{n,t-1} \right] \right\|_q \\
\leq \sum_{s=-\infty}^{-1} \left\| X^{s(j)}_{n,t+s} - \mathbb{E} \left[ X^{s(j)}_{n,t+s} | \mathcal{F}_{n,t} \right] \right\|_q + \sum_{s=-\infty}^{-1} \left\| X^{s(j)}_{n,t+s} - \mathbb{E} \left[ X^{s(j)}_{n,t+s} | \mathcal{F}_{n,t-1} \right] \right\|_q \\
\leq \sum_{s=1}^{\infty} \left\| X^{s(j)}_{n,t-s} - \mathbb{E} \left[ X^{s(j)}_{n,t-s} | \mathcal{F}_{n,t} \right] \right\|_q + \sum_{s=1}^{\infty} \left\| X^{s(j)}_{n,1-t-s} - \mathbb{E} \left[ X^{s(j)}_{n,1-t-s} | \mathcal{F}_{n,t-1} \right] \right\|_q \\
\leq \bar{c}_n k_n^{-1/2} \left( \sum_{s=1}^{\infty} \psi_{s+1} + \sum_{s=1}^{\infty} \psi_{s-1} \right) \\
\leq 2\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s < \infty.
\]

The assertion of part (a) then follows from (A.69) and (A.70).

(b) From (A.69) and (A.70), we deduce that \( \|X^{s(j)}_{n,t}\| \leq 4\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s. \) It remains to verify that \( \mathbb{E} \left[ X^*_n | \mathcal{F}_{n,t-1} \right] = 0. \) To this end, we set

\[
X^*_n(t) = \sum_{s=-m}^{m} \left\{ \mathbb{E} \left[ X_{n,t+s} | \mathcal{F}_{n,t} \right] - \mathbb{E} \left[ X_{n,t+s} | \mathcal{F}_{n,t-1} \right] \right\}.
\]

It is easy to see that \( \mathbb{E} \left[ X^*_n(t) | \mathcal{F}_{n,t-1} \right] = 0. \) We note that

\[
\left| X^{s(j)}_{n,t} \right| \leq \sum_{s=-\infty}^{\infty} \left| \mathbb{E} \left[ X^{s(j)}_{n,t+s} | \mathcal{F}_{n,t} \right] - \mathbb{E} \left[ X^{s(j)}_{n,t+s} | \mathcal{F}_{n,t-1} \right] \right|.
\]
where the right-hand side of the above display is integrable by the calculations in part (a). Since 
\( \lim_{m \to \infty} X_{n,t}^{(j)} (m) = X_{n,t}^{(j)} \) almost surely by part (a), we deduce \( \mathbb{E} \left[ X_{n,t}^{(j)} | \mathcal{F}_{n,t-1} \right] = 0 \) by using the

dominated convergence theorem.

(c) The assertion of part (c) follows from (3.1) directly. Indeed,

\[
\sum_{s=0}^{\infty} \left\| \mathbb{E} \left[ X_{n,t+s} | \mathcal{F}_{n,t-1} \right] \right\|_q \leq \tilde{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s < \infty \quad (A.71)
\]

and

\[
\sum_{s=0}^{\infty} \left\| X_{t-s-1} - \mathbb{E} [X_{t-s-1} | \mathcal{F}_{n,t-1}] \right\|_q \leq \tilde{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s < \infty. \quad (A.72)
\]

(d) The assertion follows from part (c) and the triangle inequality.

(e) We verify the assertion of part (e) as follows:

\[
\overline{X}_{n,t+1} - \overline{X}_{n,t} + X_{n,t} = \sum_{s=0}^{\infty} \mathbb{E} \left[ X_{n,t+1+s} | \mathcal{F}_{n,t} \right] - \sum_{s=0}^{\infty} \left\{ X_{t-s} - \mathbb{E} [X_{t-s} | \mathcal{F}_{n,t}] \right\}
\]

\[
- \sum_{s=0}^{\infty} \mathbb{E} \left[ X_{n,t+s} | \mathcal{F}_{n,t-1} \right] + \sum_{s=0}^{\infty} \left\{ X_{t-s-1} - \mathbb{E} [X_{t-s-1} | \mathcal{F}_{n,t-1}] \right\} + X_{n,t}
\]

\[
= \sum_{s=1}^{\infty} \mathbb{E} \left[ X_{n,t+s} | \mathcal{F}_{n,t} \right] + \sum_{s=0}^{\infty} \left\{ \mathbb{E} [X_{t-s} | \mathcal{F}_{n,t}] - X_{t-s} \right\}
\]

\[
- \sum_{s=0}^{\infty} \mathbb{E} \left[ X_{n,t+s} | \mathcal{F}_{n,t-1} \right] + \sum_{s=1}^{\infty} \left\{ X_{t-s} - \mathbb{E} [X_{t-s} | \mathcal{F}_{n,t-1}] \right\} + X_{n,t}
\]

\[
= \sum_{s=0}^{\infty} \left\{ \mathbb{E} [X_{n,t+s} | \mathcal{F}_{n,t}] - \mathbb{E} [X_{n,t+s} | \mathcal{F}_{n,t-1}] \right\} + \sum_{s=1}^{\infty} \left\{ \mathbb{E} [X_{t-s} | \mathcal{F}_{n,t}] - \mathbb{E} [X_{t-s} | \mathcal{F}_{n,t-1}] \right\}
\]

\[
= \sum_{s=-\infty}^{\infty} \left\{ \mathbb{E} [X_{n,t+s} | \mathcal{F}_{n,t}] - \mathbb{E} [X_{n,t+s} | \mathcal{F}_{n,t-1}] \right\} = X_{n,t}^*. \]

(f) The assertion follows from parts (d,e), the triangle inequality and \( \sum_k \psi_k < \infty. \) \( Q.E.D. \)

PROOF OF THEOREM 4. By Theorem 1, Lemma A4(f) and the triangle inequality, there exists a sequence \( \overline{S}_n^* \) of \( m_n \)-dimensional random vectors with distribution \( \mathcal{N} (0, \Sigma_n^*) \) such that

\[
\left\| S_n - \overline{S}_n^* \right\| = O_p(m_n^{1/2} n^{1/2} + (B_n^* m_n)^{1/3} + \tilde{c}_n m_n^{1/2} k_n^{-1/2}), \quad (A.73)
\]

where \( \Sigma_n^* = \mathbb{E} [S_n^* S_n^{*\top}] \). By Lyapunov’s inequality and Lemma A4(d),

\[
\left\| S_n - \overline{S}_n^* \right\|^2 = \sum_{j=1}^{m_n} \mathbb{E} \left[ \overline{X}_{n,t}^{(j)} - \overline{X}_{n,k_n+1}^{(j)} \right]^2 \leq K \tilde{c}_n^2 m_n k_n^{-1}. \quad (A.74)
\]

By definition, \( \Sigma_n - \Sigma_n^* = \mathbb{E} [S_n S_n^{\top} - S_n^* S_n^{*\top}] \). Hence, for any \( a \in \mathbb{R}^{m_n} \),

\[
\left\| a^\top (\Sigma_n - \Sigma_n^*) \right\|^2 \leq K \left\| \mathbb{E} \left[ a^\top (S_n - S_n^*) S_n^{\top} \right] \right\|^2 + K \left\| \mathbb{E} \left[ a^\top S_n (S_n - S_n^*) \right] \right\|^2 + K \left\| \mathbb{E} \left[ a^\top (S_n - S_n^*) (S_n - S_n^*) \right] \right\|^2. \quad (A.75)
\]

18
We now bound the terms on the majorant side of (A.75). Note that
\[
\mathbb{E}\left[a^\top(S_n - S_n^*)S_n^*\right] \Sigma_n^{-1}\mathbb{E}\left[S_n(S_n - S_n^*)^\top a\right] \leq \mathbb{E}\left[a^\top(S_n - S_n^*)\right]^2,
\]
which holds because the left-hand side is the second moment of the residual obtained by projecting \(a^\top(S_n - S_n^*)\) on the random vector \(S_n\). The first term in (A.75) can thus be bounded by
\[
\left\|\mathbb{E}\left[a^\top(S_n - S_n^*)S_n^*\right]\right\|^2 \leq \lambda_{\text{max}}(\Sigma_n)\mathbb{E}\left[a^\top(S_n - S_n^*)\right]^2 \leq K \|a\|^2 \|S_n - S_n^*\|^2_2 \quad (A.76)
\]
where the second inequality is by the Cauchy–Schwarz inequality and the boundedness of \(\lambda_{\text{max}}(\Sigma_n)\).

Turning to the second term in (A.75), we use the Cauchy–Schwarz inequality to derive
\[
\left\|\mathbb{E}\left[a^\top S_n(S_n - S_n^*)^\top\right]\right\|^2 \leq \mathbb{E}\left[(a^\top S_n)^2\right] \|S_n - S_n^*\|^2_2 \leq K \|a\|^2 \|S_n - S_n^*\|^2_2. \quad (A.77)
\]

For the third term in (A.75), we observe
\[
\left\|\mathbb{E}\left[a^\top(S_n - S_n^*)(S_n - S_n^*)^\top\right]\right\|^2 \leq \|a\|^2 \left(\mathbb{E}\left[(S_n - S_n^*)(S_n - S_n^*)^\top\right]\right)^2 \leq \|a\|^2 \left(\text{Tr}\left(\mathbb{E}\left[(S_n - S_n^*)(S_n - S_n^*)^\top\right]\right)\right)^2 \leq \|a\|^2 \|S_n - S_n^*\|^4. \quad (A.78)
\]

Combining (A.75)–(A.78), we deduce that
\[
\sup_{\|a\|=1} a^\top (\Sigma_n - \Sigma_n^*) (\Sigma_n - \Sigma_n^*)^\top a \leq K \|S_n - S_n^*\|^2_2 + K \|S_n - S_n^*\|^4_2 = O_p(c_n^2m_nk_n^{-1} + c_n^4m_n^2k_n^{-2}).
\]

Hence,
\[
\|\Sigma_n - \Sigma_n^\ast\|_S = O_p(c_n m_n^{1/2} k_n^{-1/2} + c_n^2 m_n k_n^{-1}). \quad (A.79)
\]

Let \(\bar{S}_n \equiv (\Sigma_n)^{1/2}(\Sigma_n^*)^{-1/2}\bar{S}_n^\ast\), so \(\bar{S}_n \sim \mathcal{N}(0, \Sigma_n)\). By definition,
\[
\bar{S}_n - \bar{S}_n^\ast = \left[(\Sigma_n)^{1/2} - (\Sigma_n^*)^{1/2}\right] (\Sigma_n^*)^{-1/2}\bar{S}_n^\ast
\]

which implies that
\[
\mathbb{E}\left[\|\bar{S}_n - \bar{S}_n^\ast\|^2\right] \leq \|\Sigma_n^{1/2} - (\Sigma_n^*)^{1/2}\|^2 S \mathbb{E}\left[\bar{S}_n^\ast\right] (\Sigma_n^*)^{-1} \bar{S}_n^\ast \leq K \|\Sigma_n - \Sigma_n^\ast\|^2 S \mathbb{E}\left[\bar{S}_n^\ast\right] (\Sigma_n^*)^{-1} \bar{S}_n^\ast \leq O(c_n^2m_n^2k_n^{-1} + c_n^4m_n^3k_n^{-2}) \quad (A.80)
\]

where the second inequality is by Exercise 7.2.18 in Horn and Johnson (1990) (also see Lemma A.2 in Belloni, Chernozhukov, Chetverikov, and Kato (2015)) and \(\lambda_{\text{min}}(\Sigma_n^*)^{-1} = O(1)\), and the last line follows from \(\mathbb{E}[\bar{S}_n^\ast (\Sigma_n^*)^{-1} \bar{S}_n^\ast] = m_n\) and (A.79). Hence,
\[
\left\|\bar{S}_n - \bar{S}_n^\ast\right\| = O_p(c_n m_n k_n^{-1/2} + c_n^2 m_n^{3/2} k_n^{-1}). \quad (A.81)
\]

The assertion of the theorem then follows from (A.73) and (A.81). \(Q.E.D.\)
S.A.5 Proof of Theorem 5 and Theorem 6

Lemma A5. Let $\Gamma_{X,n}(s) \equiv \mathbb{E}[X_{n,t}^{(k)} X_{n,t+s}^{(l)}]$. Under Assumption 5 and Assumption 7(iv),

$$\max \left\{ \sum_{s=-\infty}^{\infty} |s|^2 |\Gamma_{X,n}(s)| \right\} \leq K \bar{c}^2 \, k^{-1}. \tag{A.82}$$

Proof of Lemma A5. For each $s \geq 0$,

$$|\Gamma_{X,n}(s)| = \left| \mathbb{E}[X_{n,t}^{(k)} \mathbb{E}[X_{n,t+s}^{(l)}/\mathcal{F}_{n,t}]] \right| \leq \left\| X_{n,t}^{(k)} \right\|_2 \left\| \mathbb{E}[X_{n,t+s}^{(l)}/\mathcal{F}_{n,t}] \right\|_2 \leq \left\| \mathbb{E}[X_{n,t}^{(k)}/\mathcal{F}_{n,t}] \right\|_q \left\| \mathbb{E}[X_{n,t+s}^{(l)}/\mathcal{F}_{n,t}] \right\|_q \leq \psi_0 \psi_s \bar{c}^2 \, k^{-1}. \tag{A.83}$$

where the first equality is by repeated conditioning; the first inequality is by the Cauchy–Schwarz inequality; the second inequality follows from Lyapunov’s inequality; the last line is due to Assumption 5. Hence,

$$\sum_{s=-\infty}^{\infty} |s|^2 |\Gamma_{X,n}(s)| \leq 2 \sum_{s=0}^{\infty} s^2 |\Gamma_{X,n}(s)| \leq \left( 2\psi_0 \sum_{s=0}^{\infty} |s|^2 \, \psi_s \right) \bar{c}^2 \, k^{-1}. \tag{A.84}$$

By Assumption 7(iv), $K = 2\psi_0 \sum_{s=0}^{\infty} |s|^2 \, \psi_s$ is finite. This finishes the proof. Q.E.D.

Lemma A6. Under Assumptions 5, 6 and 7, we have for any $s$ with $|s| \leq k-1$,

$$\max_{1 \leq k,l \leq m_n} \left\| \frac{\Gamma_{X,n}(s)}{\mathbb{E}[\Gamma_{X,n}(s)]} \right\|_2^2 \leq K \bar{c}^4 \, k^{-1} \tag{A.85}$$

where $K > 0$ is a finite constant that does not depend on $s$.

Proof of Lemma A6. Step 1. In this step, we derive some preliminary estimates. Let $\eta_{t,s} = X_{n,t}^{(l)} X_{n,t+s}^{(k)} - \mathbb{E}[X_{n,t}^{(l)} X_{n,t+s}^{(k)}]$. We shall show that

$$|\mathbb{E}[\eta_{t,s}^2]| \leq \begin{cases} \psi_{h-s} \bar{c}^4 \, k^{-2} & \text{when } h \leq s \geq 0, \\
K (\psi_{s-h} + \psi_h^2) \bar{c}^4 \, k^{-2} & \text{when } s > h \geq 0. \end{cases} \tag{A.86}$$

We start with the case $h \geq s \geq 0$. By Assumption 7(iii), we have for all $s \geq 0$,

$$\max_{1 \leq l,k \leq m_n} \mathbb{E}\left[ \eta_{t,s}^2 \right] \leq \max_{1 \leq l,k \leq m_n} \mathbb{E}\left[ \left| X_{n,t}^{(l)} X_{n,t+s}^{(k)} \right|^2 \right] \leq \bar{c}^4 \, k^{-2}. \tag{A.87}$$

By the Cauchy–Schwarz inequality, Assumptions 7(ii) and (A.87), we deduce

$$|\mathbb{E}[\eta_{t,s} \eta_{t+h,s}]| = |\mathbb{E}[\eta_{t,s} \mathbb{E}[\eta_{t+h,s}/\mathcal{F}_{n,t+s}]]| \leq \left\| \eta_{t,s} \right\|_2 \left\| \mathbb{E}[\eta_{t+h,s}/\mathcal{F}_{n,t+s}] \right\|_2 \leq \psi_{h-s} \bar{c}^4 \, k^{-2}, \tag{A.88}$$

20
as asserted.

Turning to the case with \( s > h \geq 0 \), we first note that by the definition of \( \eta_{t,s} \) and the triangle inequality,
\[
|E[\eta_{t,s} \eta_{t+h,s}]| \leq |E[\eta_{t,h} \eta_{t+s,h}]| + |\Gamma_{n,X}(h)\left| + |\Gamma_{n,X}(s)\right|^2. \tag{A.89}
\]
By swapping \( s \) and \( h \) in (A.88), we obtain \( |E[\eta_{t,h} \eta_{t+s,h}]| \leq \psi_s c_n^4 k_n^{-2} \). By (A.83),
\[
\left|\Gamma_{n,X}(h)\right|^2 + \left|\Gamma_{n,X}(s)\right|^2 \leq K (\psi_n^2 + \psi_s^2) c_n^4 k_n^{-2}.
\]
The second claim in (A.86) then readily follows from these estimates.

**Step 2.** We now prove (A.85). Since \( \Gamma_{X,n}(s) = \Gamma_{X,n}(-s) \), it suffices to consider \( s \geq 0 \). With \( \eta_{t,s} \) defined in step 1, we can rewrite \( \Gamma_{X,n}(s) - E[\Gamma_{X,n}(s)] = \sum_{t=1}^{k_n-h} \eta_{t,s} \). Hence,
\[
\left\| \Gamma_{X,n}(s) - E[\Gamma_{X,n}(s)] \right\|_2^2 = E \left[ \left( \sum_{t=1}^{k_n-s-h} \eta_{t,s} \right)^2 \right]
\]
\[
\leq 2 \sum_{h=0}^{k_n-s-1} \sum_{t=1}^{k_n-s-h} |E[\eta_{t,s} \eta_{t+h,s}]| = 2 (R_{1,n} + R_{2,n}), \tag{A.90}
\]
where (sums over empty sets are set to zero by convention)
\[
R_{1,n} = \sum_{h=s}^{k_n-s-1} \sum_{t=1}^{k_n-s-h} |E[\eta_{t,s} \eta_{t+h,s}]|, \quad R_{2,n} = \sum_{h=0}^{(k_n-s-1) \wedge (s-1)} \sum_{t=1}^{k_n-s-h} |E[\eta_{t,s} \eta_{t+h,s}]|.
\]
By (A.86),
\[
R_{1,n} \leq \left( \sum_{h=s}^{k_n-s-1} \frac{k_n-s-h}{k_n} \psi_{h-s} \right) c_n^4 k_n^{-1} \leq \left( \sum_{h=0}^{\infty} \psi_h \right) c_n^4 k_n^{-1}, \tag{A.91}
\]
and similarly,
\[
R_{2,n} \leq K \left( \sum_{h=0}^{(k_n-s-1) \wedge (s-1)} \frac{k_n-s-h}{k_n} \left( \psi_{s-h} + \psi_h^2 + \psi_s^2 \right) \right) c_n^4 k_n^{-1}
\]
\[
\leq K \left( \sum_{h=0}^{\infty} \left( \psi_h + \psi_h^2 \right) \right) c_n^4 k_n^{-1}. \tag{A.92}
\]
Combining (A.90), (A.91) and (A.92), we deduce
\[
\left\| \Gamma_{X,n}(s) - E[\Gamma_{X,n}(s)] \right\|_2^2 \leq K \left( \sum_{h=0}^{\infty} \psi_h + \psi_s^2 \right) c_n^4 k_n^{-1}.
\]
Since \( \sum_{h=0}^{\infty} \psi_h < \infty \) and \( \sup_{s \geq 0} s \psi_s^2 < \infty \) under Assumption 5 and Assumption 7(iv), the assertion of the lemma follows from the above inequality.

\[Q.E.D.]\]
Proof of Theorem 5. Recall that \( \Gamma_{X,n}(s) \equiv \mathbb{E}[X_{n,t}X_{n,t+s}^\top] \). By definition, we can decompose

\[
\tilde{\Sigma}_n - \Sigma_n = \sum_{s=-k_n+1}^{k_n-1} \mathcal{K}(s/M_n) \left( \tilde{\Gamma}_{X,n}(s) - \mathbb{E} \left[ \tilde{\Gamma}_{X,n}(s) \right] \right) + \sum_{s=-k_n+1}^{k_n-1} (\mathcal{K}(s/M_n) - 1)(k_n - s)\Gamma_{X,n}(s). \tag{A.93}
\]

To bound the first term on the right-hand side of (A.93), we note, by the triangle inequality,

\[
\left\| \sum_{s=-k_n+1}^{k_n-1} \mathcal{K}(s/M_n) \left( \tilde{\Gamma}_{X,n}(s) - \mathbb{E} \left[ \tilde{\Gamma}_{X,n}(s) \right] \right) \right\| \leq \sum_{s=-k_n+1}^{k_n-1} |\mathcal{K}(s/M_n)| \left\| \tilde{\Gamma}_{X,n}(s) - \mathbb{E} \left[ \tilde{\Gamma}_{X,n}(s) \right] \right\| . \tag{A.94}
\]

By (A.85),

\[
\mathbb{E} \left[ \left\| \tilde{\Gamma}_{X,n}(s) - \mathbb{E} \left[ \tilde{\Gamma}_{X,n}(s) \right] \right\| \right] \leq \left( \sum_{k=1}^{m_n} \sum_{l=1}^{m_n} \left\| \tilde{\Gamma}_{X,n}(s) - \mathbb{E}[\tilde{\Gamma}_{X,n}(s)] \right\|^2 \right)^{1/2} \leq K\bar{c}_n^2 m_n k_n^{-1/2}.
\]

Combining this estimate with (A.94), we deduce

\[
\mathbb{E} \left[ \left\| \sum_{s=-k_n+1}^{k_n-1} \mathcal{K}(s/M_n) \left( \tilde{\Gamma}_{X,n}(s) - \mathbb{E} \left[ \tilde{\Gamma}_{X,n}(s) \right] \right) \right\| \right] \leq K\bar{c}_n^2 m_n k_n^{-1/2} \sum_{s=-k_n+1}^{k_n-1} |\mathcal{K}(s/M_n)| \leq K\bar{c}_n^2 m_n M_n k_n^{-1/2},
\]

where the second inequality follows from Assumption 6. From here, we deduce

\[
\sum_{s=-k_n+1}^{k_n-1} \mathcal{K}(s/M_n) \left( \tilde{\Gamma}_{X,n}(s) - \mathbb{E} \left[ \tilde{\Gamma}_{X,n}(s) \right] \right) = O_p(\bar{c}_n^2 m_n M_n k_n^{-1/2}). \tag{A.95}
\]

We now turn to the second term on the right-hand side of (A.93). By definition,

\[
\left\| \sum_{s=-k_n+1}^{k_n-1} (\mathcal{K}(s/M_n) - 1)(k_n - s)\Gamma_{X,n}(s) \right\|^2 = \sum_{k=1}^{m_n} \sum_{l=1}^{m_n} \left\| \sum_{s=-k_n+1}^{k_n-1} (\mathcal{K}(s/M_n) - 1)(k_n - s)\Gamma^{(k,l)}_{X,n}(s) \right\|^2 . \tag{A.96}
\]

Let \( r = r_1 \wedge r_2 \). By Assumption 6, we can fix some (small) constant \( \varepsilon \in (0, 1) \) such that

\[
\frac{|1 - \mathcal{K}(x)|}{|x|^p} \leq \frac{|1 - \mathcal{K}(x)|}{|x|^{p_1}} \leq K \text{ for } x \in [-\varepsilon, \varepsilon]. \tag{A.97}
\]
By the triangle inequality

\[
\left| \sum_{s = -k_n + 1}^{k_n - 1} (K(s/M_n) - 1)(k_n - s)\Gamma_{X,n}^{(k,l)}(s) \right| \\
\leq k_n M_n^{-r} \sum_{|s| \leq \varepsilon M_n} \left| \frac{K(s/M_n) - 1}{|s/M_n|^r} \right| |s|^r \left| \Gamma_{X,n}^{(k,l)}(s) \right| \\
+ k_n \sum_{\varepsilon M_n < |s| < k_n} |K(s/M_n) - 1| \left| \Gamma_{X,n}^{(k,l)}(s) \right|.
\]  

(A.98)

By (A.97),

\[
\sum_{|s| \leq \varepsilon M_n} \left| \frac{K(s/M_n) - 1}{|s/M_n|^r} \right| |s|^r \left| \Gamma_{X,n}^{(k,l)}(s) \right| \leq K \sum_{|s| \leq \varepsilon M_n} |s|^r \left| \Gamma_{X,n}^{(k,l)}(s) \right|
\]

\[
\leq K \sum_{s = -\infty}^{\infty} |s|^r \left| \Gamma_{X,n}^{(k,l)}(s) \right|.
\]  

(A.99)

Since \( K(\cdot) \) is bounded (Assumption 6),

\[
\sum_{\varepsilon M_n < |s| < k_n} |K(s/M_n) - 1| \left| \Gamma_{X,n}^{(k,l)}(s) \right| \leq K \sum_{\varepsilon M_n < |s| < k_n} \left| \Gamma_{X,n}^{(k,l)}(s) \right|
\]

\[
\leq K M_n^{-r} \sum_{s = -\infty}^{\infty} |s|^r \left| \Gamma_{X,n}^{(k,l)}(s) \right|.
\]  

(A.100)

Combining (A.98), (A.99) and (A.100), we deduce

\[
\left| \sum_{s = -k_n + 1}^{k_n - 1} (K(s/M_n) - 1)(k_n - s)\Gamma_{X,n}^{(k,l)}(s) \right| \\
\leq K k_n M_n^{-r} \sum_{s = -\infty}^{\infty} |s|^r \left| \Gamma_{X,n}^{(k,l)}(s) \right| \leq K \tilde{c}_n^2 M_n^{-r}
\]  

(A.101)

where the second inequality is by (A.82). By (A.96) and (A.101),

\[
\sum_{s = -k_n + 1}^{k_n - 1} (K(s/M_n) - 1)(k_n - s)\Gamma_{X,n}(s) = O(\tilde{c}_n^2 m_n M_n^{-r}).
\]  

(A.102)

The assertion of the theorem then follows from (A.93), (A.95) and (A.102).

Q.E.D.

**Proof of Theorem 6.** By Theorem 5,

\[
\| \tilde{\Sigma}_n - \Sigma_n \| = O_p(\tilde{c}_n^2 m_n(M_n k_n^{-1/2} + M_n^{-r_1/r_2})).
\]  

(A.103)

To prove the assertion of the theorem, it remains to show that

\[
\| \tilde{\Sigma}_n - \bar{\Sigma}_n \| = O_p(M_n m_n^{1/2} \delta_{\theta,n}).
\]  

(A.104)
By the definitions of $\hat{\Gamma}_{X,n}(s)$ and $\tilde{\Gamma}_{X,n}(s)$, for any $s \geq 0$, we can decompose
\[
\hat{\Gamma}_{X,n}(s) - \tilde{\Gamma}_{X,n}(s) = k^{-1}_n \sum_{t=1}^{k_n - s} \left[ g(Z_t, \hat{\theta}_n) - g(Z_t, \theta_0) \right] \left[ g(Z_{t+s}, \hat{\theta}_n) - g(Z_{t+s}, \theta_0) \right]^\top 
+ k^{-1}_n \sum_{t=1}^{k_n - s} \left[ g(Z_t, \hat{\theta}_n) - g(Z_t, \theta_0) \right] g(Z_{t+s}, \theta_0)^\top 
+ k^{-1}_n \sum_{t=1}^{k_n - s} g(Z_t, \theta_0) \left[ g(Z_{t+s}, \hat{\theta}_n) - g(Z_{t+s}, \theta_0) \right]^\top .
\] (A.105)

Therefore, by the triangle inequality and the Cauchy–Schwarz inequality,
\[
\max_{|s| \leq k_n - 1} \left\| \hat{\Gamma}_{X,n}(s) - \tilde{\Gamma}_{X,n}(s) \right\| 
\leq k^{-1}_n \sum_{t=1}^{k_n} \left\| g(Z_t, \hat{\theta}_n) - g(Z_t, \theta_0) \right\|^2 
+ 2 \left( k^{-1}_n \sum_{t=1}^{k_n} \left\| g(Z_t, \hat{\theta}_n) - g(Z_t, \theta_0) \right\| \right)^1/2 \left( k^{-1}_n \sum_{t=1}^{k_n} \left\| g(Z_t, \theta_0) \right\| \right)^1/2 .
\] (A.106)

By Assumption 8(ii) and Markov’s inequality,
\[
k^{-1}_n \sum_{t=1}^{k_n} \left\| g(Z_t, \theta_0) \right\|^2 = O_p(M_n). 
\] (A.107)

By Assumption 8(i), (A.106) and (A.107), we deduce
\[
\max_{|s| \leq k_n - 1} \left\| \hat{\Gamma}_{X,n}(s) - \tilde{\Gamma}_{X,n}(s) \right\| = O_p(m_n^{1/2} \delta_{\theta,n}).
\] (A.108)

By the triangle inequality, (A.108) and Assumption 6(i), we deduce
\[
\left\| \hat{\Sigma}_n - \tilde{\Sigma}_n \right\| \leq \sum_{s=-k_n+1}^{k_n-1} \left| \mathcal{K}(s/M_n) \right| \left\| \hat{\Gamma}_{X,n}(s) - \tilde{\Gamma}_{X,n}(s) \right\| 
\leq \max_{|s| \leq k_n - 1} \left\| \hat{\Gamma}_{X,n}(s) - \tilde{\Gamma}_{X,n}(s) \right\| \sum_{s=-k_n+1}^{k_n-1} \left| \mathcal{K}(s/M_n) \right| 
= O_p(M_n m_n^{1/2} \delta_{\theta,n})
\] (A.109)

as claimed in (A.104). This finishes the proof. 

Q.E.D.
S.A.6 Technical derivations of the conditional moment restriction in the search and matching model

In this appendix, we derive the conditional moment restriction (4.5) in the main text. Recall that the equilibrium is characterized by the following Bellman equations:

\[ J_p = p - w_p + \delta (1 - s) \mathbb{E}_p [J_{p'}], \tag{A.110} \]
\[ V_p = -c_p + \delta q (\theta_p) \mathbb{E}_p [J_{p'}], \tag{A.111} \]
\[ U_p = z + \delta \left\{ f (\theta_p) \mathbb{E}_p [W_{p'}] + (1 - f (\theta_p)) \mathbb{E}_p [U_{p'}] \right\}, \tag{A.112} \]
\[ W_p = w_p + \delta \left\{ (1 - s) \mathbb{E}_p [W_{p'}] + s \mathbb{E}_p [U_{p'}] \right\}, \tag{A.113} \]

the free entry condition \( V_p = 0 \) and the Nash bargaining solution

\[ J_p = (W_p - U_p) (1 - \beta)/\beta. \tag{A.114} \]

Taking a difference between (A.112) and (A.113) yields

\[ W_p - U_p = w_p - z + \delta (1 - s - f (\theta_p)) \mathbb{E}_p [W_{p'} - U_{p'}]. \tag{A.115} \]

Combining (A.115) with (A.114), we derive

\[ J_p = \frac{1 - \beta}{\beta} (w_p - z) + \delta (1 - s - f (\theta_p)) \mathbb{E}_p [J_{p'}]. \tag{A.116} \]

From (A.110) and (A.116), we can solve for the wage function

\[ w_p = \beta p + (1 - \beta) z + \beta \delta f (\theta_p) \mathbb{E}_p [J_{p'}]. \tag{A.117} \]

Note that the free entry condition implies

\[ \delta q (\theta_p) \mathbb{E}_p [J_{p'}] - c_p = 0. \tag{A.118} \]

Since \( f (\theta) / q (\theta) = \theta \), we can rewrite (A.117) as

\[ w_p = \beta p + (1 - \beta) z + \beta \theta_p c_p. \tag{A.119} \]

We can rewrite (A.118) as \( \delta \mathbb{E}_p [J_{p'}] = c_p / q (\theta_p) \). Plugging this and (A.119) into (A.110), we deduce

\[ J_p = (1 - \beta) (p - z) - \beta \theta_p c_p + (1 - s) \frac{c_p}{q (\theta_p)}. \tag{A.120} \]

Finally, plugging (A.120) into (A.118) yields

\[ \delta q (\theta_p) \mathbb{E}_p \left[ (1 - \beta) (p' - z) - \beta \theta_{p'} c_{p'} + (1 - s) \frac{c_{p'}}{q (\theta_{p'})} \right] - c_p = 0. \tag{A.121} \]
In standard calibration analysis, one can solve $\theta_p$ from this equation, and then calibrate parameters by matching certain model-implied quantities (e.g., the average market tightness, the job finding rate, etc.) with their empirical counterparts.

From an econometric viewpoint, we consider (A.121) alternatively as a conditional moment restriction on observed data. Replacing $p$ and $\theta$ with their observed time series yields

$$
\delta q (\theta) \mathbb{E}_t \left[ (1 - \beta) (p_{t+1} - z) - \beta \theta_{t+1} c_{t+1} + (1 - s) \frac{c_{t+1}}{q(\theta_{t+1})} \right] - c_t = 0,
$$

(A.122)

where we write $c_t$ in place of $c_p$ (recall (A.118)) and use $\mathbb{E}_t$ to denote the conditional expectation given the time-$t$ information.\(^1\) The conditional moment restriction (4.5) is then obtained from (A.122).

S.B Additional technical results

This appendix collects additional technical results. Section S.B.1 verifies Assumption 1(ii) under primitive sufficient conditions. Section S.B.2 provides an explicit rate of the strong approximation. Section S.B.3 provides examples for the strong approximation result under primitive conditions. Section S.B.4 provides primitive conditions for Assumption 2 in the main text.

S.B.1 Sufficient conditions for Assumption 1(ii)

We illustrate how to verify Assumption 1(ii) in the following proposition. The primitive conditions mainly require that the volatility $V_{n,t}$ is weakly dependent, here formalized in terms of strong and uniform mixing coefficients.

**Proposition B1.** Suppose (i) $V_{n,t} = v_{n,t}/k_n$ for some process $(v_{n,t})_{t\geq 0}$ taking values in $\mathbb{R}^{m_n \otimes m_n}$ such that $\sup_{t,j,l} \|v_{n,t}^{(j,l)}\|_q \leq \bar{c}_n^2$ for some constant $q \geq 2$ and some real sequence $\bar{c}_n$; either (ii) $q > 2$ and $v_{n,t}$ is strong mixing with mixing coefficient $\alpha_k$ satisfying $\sum_{k=1}^{k_n} \alpha_k^{1-2/q} < K$, or (iii) $q = 2$ and $v_{n,t}$ is uniform mixing with mixing coefficient $\phi_k$ satisfying $\sum_{k=1}^{k_n} \phi_k^{1/2} < K$. Then, uniformly for all sequence $h_n$ that satisfies $h_n \leq k_n$,

$$
\left\| \sum_{t=1}^{h_n} (V_{n,t} - \mathbb{E}[V_{n,t}]) \right\|_2 = O(r_n), \quad \text{for} \quad r_n \equiv \bar{c}_n^2 m_n k_n^{-1/2}.
$$

(B.1)

Consequently, condition (ii) of Assumption 1 holds provided that $r_n = o(1)$.

**COMMENT.** The sequence $\bar{c}_n$ bounds the magnitude of the $k_n^{1/2}X_{n,t}$ array. It is instructive to illustrate the “typical” magnitude of $\bar{c}_n$ in the context of series estimation, where $X_{n,t}$ is the score

\(^1\)Since the state process is Markovian, the time-$t$ information set is spanned by $p_t$, that is, $\mathbb{E}_t [\cdot] = \mathbb{E} [\cdot | p_t]$. 

26
process given by \( X_{n,t} = u_t P(X_t)k_n^{-1/2} \). Suppose the \( F_{n,t-1} \)-conditional 2\( q \)th moment of \( u_t \) is uniformly bounded. Then
\[
\|v_t^{(j,l)}\|_q \leq C \|p_j(X_t)p_t(X_t)\|_q.
\]
Therefore, \( \bar{c}_n = O(m_n^{1/2}) \) if \( P(\cdot) \) collects Legendre polynomials, and \( \bar{c}_n = O(1) \) if \( P(\cdot) \) consists of trigonometric polynomials or splines. In these cases, \( r_n = o(1) \) is implied by \( m_n \ll k_n^{1/4} \) and \( m_n \ll k_n^{1/2} \), respectively.

**Proof of Proposition B1.** We observe that
\[
\mathbb{E} \left[ \sum_{t=1}^{h_n} \left( V_{n,t}^{(j,l)} - \mathbb{E}[V_{n,t}^{(j,l)}] \right)^2 \right] = \frac{1}{k_n^2} \mathbb{E} \left[ \sum_{t=1}^{h_n} \left( v_{n,t}^{(j,l)} - \mathbb{E}[v_{n,t}^{(j,l)}] \right)^2 \right] \tag{B.2}
\]
\[
= \frac{1}{k_n^2} \sum_{s,t=1}^{h_n} \mathbb{E} \left[ \left( v_{n,t}^{(j,l)} - \mathbb{E}[v_{n,t}^{(j,l)}] \right) \left( v_{n,s}^{(j,l)} - \mathbb{E}[v_{n,s}^{(j,l)}] \right) \right] \]
\[
\leq \frac{2}{k_n^2} \sum_{k=0}^{h_n-1} \sum_{t=k+1}^{h_n} \mathbb{E} \left[ \left( v_{n,t}^{(j,l)} - \mathbb{E}[v_{n,t}^{(j,l)}] \right) \left( v_{n,t-k}^{(j,l)} - \mathbb{E}[v_{n,t-k}^{(j,l)}] \right) \right] .
\]

We then prove the assertions for the \( \alpha \)-mixing and \( \phi \)-mixing cases separately.

**The \( \alpha \)-mixing case.** By the covariance inequality for strong mixing processes (see, e.g., Corollary 14.3 of Davidson (1994)),
\[
\left| \mathbb{E} \left[ \left( v_{n,t}^{(j,l)} - \mathbb{E}[v_{n,t}^{(j,l)}] \right) \left( v_{n,s}^{(j,l)} - \mathbb{E}[v_{n,s}^{(j,l)}] \right) \right] \right| \leq K \alpha_{k}^{-2/q} \| v_{n,t}^{(j,l)} \|_q \| v_{n,t-k}^{(j,l)} \|_q .
\]

Therefore, we can further bound the terms in (B.2) as follows
\[
\mathbb{E} \left[ \sum_{t=1}^{h_n} \left( V_{n,t}^{(j,l)} - \mathbb{E}[V_{n,t}^{(j,l)}] \right)^2 \right] \leq K \frac{c_n^2}{k_n^2} \sum_{k=0}^{h_n-1} (h_n - k) \alpha_{k}^{-2/q} \leq K \bar{c}_n^2 k_n^{-1} ,
\]
where the second inequality is due to condition (ii). The assertion of the proposition readily follows.

**The \( \phi \)-mixing case.** The case with uniform mixing can be proved similarly. Indeed, by the covariance inequality for uniform mixing processes (see, e.g., Corollary 14.5 of Davidson (1994)) and condition (iii),
\[
\mathbb{E} \left[ \sum_{t=1}^{h_n} \left( V_{n,t}^{(j,l)} - \mathbb{E}[V_{n,t}^{(j,l)}] \right)^2 \right] \leq K \frac{c_n^2}{k_n^2} \sum_{k=0}^{h_n-1} (h_n - k) \phi_{k}^{1/2} \left( \sup_{j,l,t} \| v_{n,t}^{(j,l)} \|_2 \right)^2 \leq K \bar{c}_n^2 k_n^{-1} .
\]

From here, the assertion of the proposition for the \( \phi \)-mixing case readily follows.
S.B.2 Convergence rate of strong approximation

The strong approximation rate in (2.3) can be simplified under additional—but mild—assumptions. Corollary B1, below, provides a pedagogical example of this kind. We remind the reader that the typical order of each component of the (normalized) variable $X_{n,t}$ is $k_n^{-1/2}$, and hence it is reasonable to assume that its fourth moment is of order $k_n^{-2}$.

**Corollary B1.** Under the same setting as Theorem 1, if $\sup_{t,j} E[(X_{n,t}^{(j)})^4] = O(k_n^{-2})$ holds in addition, then $B_n = O(k_n^{-1/2}m_n^{3/2})$. Consequently, $\|S_n - \tilde{S}_n\| = O_p(m_n^{1/2}r_n^{1/2} + m_n^{5/6}k_n^{-1/6})$.

**Comment.** This corollary suggests that $m_n \ll k_n^{1/5}$ is needed for the validity of the strong approximation. The dimension $m_n$ thus cannot grow too fast relative to the sample size $k_n$.

**Proof of Corollary B1.** We can bound $B_n = \sum_{t=1}^{k_n} E[\|X_{n,t}\|^3]$ as follows:

$$
\sum_{t=1}^{k_n} E[\|X_{n,t}\|^3] \leq \sum_{t=1}^{k_n} \left( E[\|X_{n,t}\|^4] \right)^{3/4} = \sum_{t=1}^{k_n} \left( \sum_{j=1}^{m_n} \sum_{l=1}^{m_n} E \left[ (X_{n,t}^{(j)})^2 (X_{n,t}^{(l)})^2 \right] \right)^{3/4} \leq \sum_{t=1}^{k_n} \left( \sum_{j=1}^{m_n} \left( E \left[ (X_{n,t}^{(j)})^4 \right] \right)^{1/2} \right)^{3/2} = O\left(k_n^{-1/2}m_n^{3/2}\right),
$$

where the first inequality is by Jensen’s inequality; the second inequality is by the Cauchy–Schwarz inequality and the last line follows from $\sup_{t,j} E[(X_{n,t}^{(j)})^4] = O(k_n^{-2})$. Plugging the estimate above into (2.3), we readily deduce the assertion of Corollary B1. Q.E.D.

S.B.3 Examples for Theorem 4 under primitive conditions

Condition (ii) of Theorem 4 in the main text is high-level in nature in that it is stated for the approximating martingale difference $X_{n,t}^*$ instead of for the underlying mixingale $X_{n,t}$ directly. In this subsection, we provide two examples so as to illustrate how to verify this high-level condition under primitive conditions. The first example concerns linear processes and is relatively simple to describe.

**Example 1 (Martingale Approximation for Linear Processes).** Let $(\varepsilon_{n,t}, F_{n,t})$ be a martingale difference array such that $\|\varepsilon_{n,t}\|_q \leq \bar{c}_n k_n^{-1/2}$ uniformly for some $q \geq 3$. Suppose that $X_{n,t}$ is a linear process with the form $X_{n,t} = \sum_{|j|<\infty} \theta_j \varepsilon_{n,t-j}$, where the coefficients $(\theta_j)$ satisfy $\sum_{|j|<\infty} |j\theta_j| < \infty$. Then $(X_{n,t})$ is an $L_r$-mixingale that satisfies Assumption 5 with
ψ_k = \sum_{|j| \geq k} |\theta_j| \quad \text{(see, e.g., Example 16.2 in Davidson (1994))}; in particular, the summability condition \( \sum_{k=0}^{\infty} \psi_k < \infty \) is implied by \( \sum_{|j| < \infty} |j\theta_j| < \infty \). In this case, the martingale difference component \( X^*_n,t \) has a closed-form expression \( X^*_n,t = (\sum_{|j| < \infty} \theta_j) \varepsilon_{n,t} \), which verifies the conditions in Theorem 4 if and only if \( \varepsilon_{n,t} \) satisfies Assumption 1. In the simple case when \( \varepsilon_{n,t} \) has constant covariance matrix \( \Sigma \), the pre-asymptotic covariance matrix of \( S_n^* \) is \( (\sum_{|j| < \infty} \theta_j)^2 \Sigma \), which is exactly the long-run covariance matrix of \( X_{n,t} \); consequently, the third error term on the right-hand side of (3.4) is absent.  

The second example, which concerns mixing-type primitive conditions, is slightly more complicated. In this example, we suppose that \( X_{n,t} = k^{-1/2} \varepsilon_t \), where \( (\varepsilon_t)_{\infty}^{\infty} \) is an \( m_n \)-dimensional zero mean strictly stationary (strong or uniform) mixing sequence with mixing coefficients \( (\varphi_s)_{s=0}^{\infty} \). Let the filtration be defined as \( \mathcal{F}_{n,t} = \sigma(\varepsilon_s : s \leq t) \). We consider the following regularity condition.

**Assumption B1.**  
(i) \( \sup_{t,j} ||\varepsilon_t(j)||_\kappa \leq c_{\kappa,n} \) where the sequence \( c_{\kappa,n} \) is bounded away from zero, \( \kappa > 5 \) for the strong mixing case and \( \kappa > 4 \) for the uniform mixing case; (ii) \( \sum_{s=0}^{\infty} \varphi_s^{(\kappa-4)/(5\kappa)} < \infty \) for the strong mixing case and \( \sum_{s=0}^{\infty} \varphi_s^{1/2} < \infty \) for the uniform mixing case; (iii) the eigenvalues of \( \Sigma_n \equiv \mathbb{E}[S_nS_n^\top] \) are bounded from above and away from zero; and (iv) \( c_{\kappa,n} m_n^{5/6} k_n^{-1/6} = o(1) \).

Assumption B1(i) imposes uniform moment bounds on \( (\varepsilon_t)_{\infty}^{\infty} \). Assumption B1(ii) restricts the level of dependence. Assumption B1(iii) requires that the covariance matrix \( \Sigma_n \) is non-degenerate. Assumption B1(iv) mainly restricts the rate at which the dimension of \( \varepsilon_t \) grows to infinity. Under this assumption, we can verify the conditions in Theorem 4 and obtain a strong approximation for \( S_n \), as stated by the following proposition.

**Proposition B2.** Under Assumption B1, we have

\[
\|S_n - \tilde{S}_n\| = O_p(c_{\kappa,n} m_n^{5/6} k_n^{-1/6})
\]

where \( \tilde{S}_n \) is an \( m_n \)-dimensional random vector with distribution \( \mathcal{N}(0, \Sigma_n) \).

**Comment.** We can compare this strong approximation result with Theorem 1 of Dehling (1983). For example, assuming that the strong mixing coefficient converges to zero sufficiently fast and \( c_{\kappa,n} = O(1) \), (1.13) in Dehling (1983) implies that the strong approximation error converges at a rate that is slower than \( m_n^{11/6} k_n^{-1/900} \) (this is the best-case scenario obtained by setting \( d = m_n \), \( \delta = 2/3 \), \( \varepsilon = 1 \) and \( \rho_{2+\delta} = m_n \) in that paper). Evidently, the \( m_n^{5/6} k_n^{-1/6} \) rate implied by Proposition B2 improves significantly the rate derived in Dehling (1983).

**Proof of Proposition B2.** Step 1. In this step, we verify the conditions of Theorem 4. Condition (iii) of Theorem 4 coincides with Assumption B1(iii). It remains to verify conditions (i) and (ii) of that theorem.
We first show that Assumption 5 holds for the $X_{n,t}$ array (i.e., condition (i) of Theorem 4). Let $q = 5\kappa/(\kappa + 1)$ and $q = 4$ for the strong and the uniform mixing case, respectively. Then by Assumption B1(i) and the mixing inequality (see, e.g., Theorem 14.2 and Theorem 14.4 in Davidson (1994)),

$$\|E\left[ X_{n,t}^{(j)} | F_{n,t-1} \right] \|_q \leq 6c_{\kappa,n}k_n^{-1/2} \varphi_s^{1/q-1/\kappa}$$

(B.3)
in the strong mixing case, and

$$\|E\left[ X_{n,t}^{(j)} | F_{n,t-1} \right] \|_q \leq 2c_{\kappa,n}k_n^{-1/2} \varphi_s^{1-1/\kappa}$$

(B.4)
in the uniform mixing case. Therefore, $(X_{n,t})_{t=-\infty}^{\infty}$ is an $L_q$-mixingale array with $c_n = 6c_{\kappa,n}$ and $\psi_s = \varphi_s^{1/q-1/\kappa}$ for the strong mixing case, and $c_n = 2c_{\kappa,n}$ and $\psi_s = \varphi_s^{1-1/\kappa}$ for the uniform mixing case. It remains to check the summability condition $\sum_{s=0}^{\infty} \psi_s < \infty$; this holds under Assumption B1(ii) because $1/q - 1/\kappa = (\kappa - 4)/(5\kappa)$ for the strong mixing case and $1 - 1/\kappa > 1/2$ for the uniform mixing case.

We now verify condition (ii) of Theorem 4, that is, the approximating martingale difference $X_{n,t}^*$ satisfies Assumption 1. Note that $X_{n,t}^* = k_n^{-1/2} \varepsilon_t^*$ where

$$\varepsilon_t^* \equiv \sum_{s=-\infty}^{\infty} \{E[\varepsilon_{t+s}|F_{n,t}] - E[\varepsilon_{t+s}|F_{n,t-1}]\}.$$  

(B.5)

We denote the conditional covariance matrix of $X_{n,t}^*$ by

$$V_{n,t}^* = E\left[ X_{n,t}^* X_{n,t}^{*\top} | F_{n,t-1} \right] = k_n^{-1} v_t^*,$$

where $v_t^* \equiv E[\varepsilon_t^*\varepsilon_t^{*\top} | F_{n,t-1}]$. Since $\varepsilon_t$ is stationary, $(\varepsilon_t^*)_{t \geq 1}$ is also stationary. In particular,

$$k_n E[V_{n,t}^*] = E[v_t^*] = \Sigma_n^*.$$  

(B.6)

Like (A.79), we can show that

$$\|\Sigma_n - \Sigma_n^*\|_S = O_p(c_{\kappa,n} m_n^{1/2} k_n^{-1/2} + c_{\kappa,n}^2 m_n k_n^{-1}) = o(1),$$

(B.7)

where the second equality is due to Assumption B1(iv). Hence, Assumption B1(iii) implies that the eigenvalues of $\Sigma_n^*$ is bounded away from zero and from above. In view of (B.6), we see that $X_{n,t}^*$ satisfies Assumption 1(i). Finally, we can verify that $X_{n,t}^*$ satisfies Assumption 1(ii) by using Proposition B1, with

$$r_n = c_{\kappa,n}^2 m_n k_n^{-1/2}.$$  

(B.8)

**Step 2.** By the derivations in step 1, we can apply Theorem 4 to show that

$$\left\| S_n - \widetilde{S}_n \right\| = O_p(c_{\kappa,n} m_n^{1/2} k_n^{-1/2}) + O_p(m_n^{1/2} t_n^{1/2} + (B_n^* m_n)^{1/3}) + O_p(c_{\kappa,n} m_n k_n^{-1/2} + c_{\kappa,n}^2 m_n^{3/2} k_n^{-1}).$$
Following the same argument as in Corollary B1, we deduce $B^*_n = O(c_{n,m}^3m_{n,k}^{-1/2})$. Using this estimate and (B.8), we can simplify the error bound above as 

$$
\|S_n - \tilde{S}_n\| = O_p(c_{n,m}5/6m_{n,k}^{-1/6} + c_{n,m}m_{n,k}^{-1/4} + c_{n,m}3/2m_{n,k}^{-1})
$$

Under the maintained assumptions, $m_n \ll k_{n}^{1/5}$ and $c_{n,m} \ll k_{n}^{-5/6}$, which further imply that $\|S_n - \tilde{S}_n\| = O_p(c_{n,m}5/6m_{n,k}^{-1/6})$ as asserted. Q.E.D.

S.B.4 Primitive conditions for Assumption 2

S.B.4.1 The case with martingale difference residuals

In this subsection, we illustrate how to verify Assumption 2 under the following primitive conditions.

Assumption B2. (i) $(X_t)_t$ is a strictly stationary strong mixing process with mixing coefficient $(\varphi_s)_{s=0}^\infty$ satisfying $\sum_{s=1}^\infty \varphi_s^{1-2/\kappa} \leq K$ for some finite constant $\kappa > 2$; (ii) $E[u_t|\mathcal{F}_{n,t-1}] = 0$ and $E[u_t^2|\mathcal{F}_{n,t-1}] = \sigma^2_n(X_t)$ where $\mathcal{F}_{n,t-1}$ denotes the $\sigma$-field generated by $\{(X^s_t, u_{s-1})\}_{s=1}^t$ for $t = 1, \ldots, n$; (iii) $\sigma^2_n(x)$ is continuous and is bounded from above and away from zero for any $x \in \mathcal{X}$ and for any $n$; (iv) $E[u_t^4|X_t] \leq K$ almost surely for any $t$; (v) the eigenvalues of $E[P(X_t)P(X_t)^\top]$ are bounded from above and away from zero; (vi) $\max_{1 \leq k \leq m_n} \sup_{x \in \mathcal{X}} |p_k(x)| \leq \zeta_n$ where $\zeta_n$ is a non-decreasing sequence and $\log(\zeta_n) = O(\log(m_n))$; (vii) there exist $\rho_h > 0$ and $b_n^* \in \mathbb{R}^{m_n}$ such that

$$
\sup_{x \in \mathcal{X}} \left| P(x)^\top b_n^* - h(x) \right| = O(m_n^{-\rho_h});
$$

(viii) $\inf_{x \in \mathcal{X}} \|P(x)\| \geq c$ for all $m_n$ and some constant $c > 0$; (ix) $n^{1/2}m_n^{-\rho_h} + \zeta_n^{-1/\kappa}m_n^{-1/4} + \zeta_n^{-1/3}m_n^{-5/6}n^{-1/6} = o(1)$.

Assumption B2(i) imposes restrictions on the serial dependence of the nonparametric regressor $X_t$. Assumption B2(ii) implies that the residual $u_t$ is a martingale difference. Assumption B2(iii) imposes restrictions on the conditional variance of $u_t$. Assumption B2(v) together with Assumption 2(ii), which is a standard regularity condition in the series estimation literature (see, e.g., Andrews (1991), Newey (1997), Chen (2007) and Belloni, Chernozhukov, Chetverikov, and Kato (2015)). Assumption B2(iv) imposes moment bound on the residual $u_t$, which is also standard. Assumption B2(vi) defines a uniform upper bound of the approximating functions $P(\cdot)$. Assumption B2(vii) assumes that the unknown function $h(\cdot)$ can be approximated by $P(x)^\top b_n^*$ with approximation error $O(m_n^{-\rho_h})$ under the uniform metric, where the $\rho_h$ coefficient may be further related to the level of differentiability of $h(\cdot)$ and the dimension of $X_t$ for specific basis functions, with an appeal to well-known numerical approximation theory. Assumption B2(viii) holds trivially if the basis functions include the constant function. Assumption B2(ix) specifies the growth rate of $m_n$ and imposes restriction on the series approximation error.
Proposition B3. The sufficient conditions of Theorem 2(a) hold under Assumption B2.

Proof of Proposition B3. First, \( \log(\zeta_n) = O(\log(m_n)) \) is maintained in Assumption B2(vi).

By Assumption B2(vii, viii),

\[
\sup_{x \in \mathcal{X}} \frac{n^{1/2} |h(x) - P(x)^\top b_n|}{\|P(x)\|} = O(n^{1/2}m_n^{-\rho_h}). \tag{B.9}
\]

Therefore, Assumption 2(i) holds with \( \delta_{1,n} = n^{1/2}m_n^{-\rho_h} \), where \( \delta_{1,n} = o(1) \) under Assumption B2(ix). By Assumptions B2(i, ii),

\[
Q_n = \mathbb{E} \left[ P(X_t)P(X_t)\top \right] \quad \text{and} \quad A_n = \mathbb{E} \left[ \sigma_n^2(X_t)P(X_t)P(X_t)\top \right].
\]

Assumption 2(ii) is directly implied by Assumptions B2(iii, v) in view of the above expressions. Assumptions 2(iii), 2(iv) and 2(v) have been verified in Lemma B1, Lemma B2 and Lemma B3, respectively; \( \delta_{j,n} = o(1), j \in \{2, 3, 4\} \), holds because of Assumption B2(ix). Therefore, Assumption 2 holds under Assumption B2. Moreover, in Lemma B2 and Lemma B3, we show that Assumptions 2(iv, v) hold with \( \delta_{2,n} = \delta_{4,n} = \zeta_n^{2-2/\kappa}m_n^{n-1/2} \). Therefore, by Assumption B2(ix), \( m_n^{-2}(\delta_{2,n} + \delta_{4,n}) = o(1) \).

Proposition B3 implies that Theorem 2(a) holds with

\[
\delta_n = n^{1/2}m_n^{-\rho_h} + \zeta_n^{1-1/\kappa}m_n^{n-1/4} + \zeta_n^{1/3}m_n^{5/6}m_n^{-1/6} = o(1).
\]

The uniform size control in Theorem 2(b) further requires that \( \delta_n(\log m_n)^{1/2} = o(1) \). We next discuss this restriction for some popular approximating functions.

For the splines or trigonometric series, \( \zeta_n \leq K \). In this case, Assumption B2(ix) is reduced to

\[
n^{1/2}m_n^{-\rho_h} + m_n^{n-1/4} + m_n^{5/6}m_n^{-1/6} = o(1).
\]

The above restriction requires that \( m_n = o(n^{1/5}) \) and \( \rho_h > 5/2 \). Therefore, the sufficient conditions of Theorem 2(b) hold if \( m_n = o(n^{1/5}(\log n)^{-1/2}) \) and \( \rho_h > 5/2 \). For the Legendre polynomials, \( \zeta_n \leq Km_n^{1/2} \). In this case, Assumption B2(ix) becomes

\[
n^{1/2}m_n^{-\rho_h} + m_n^{3/2-1/(2\kappa)n-1/4} + m_n^{n-1/6} = o(1).
\]

The above restriction requires that \( m_n = o(n^{1/6}) \) and \( \rho_h > 3 \). Therefore, Assumption B2(ix) and \( \delta_n(\log m_n)^{1/2} = o(1) \) hold if \( m_n = o(n^{1/6}(\log n)^{-1/2}) \) and \( \rho_h > 3 \). In the i.i.d. setting, Belloni, Chernozhukov, Chetverikov, and Kato (2015) imposes \( m_n = o(n^{1/5}(\log n)^{-2/5}a_n^{-6/5}) \) for the splines or trigonometric series and \( m_n = o(n^{1/6}(\log n)^{-1/3}a_n^{-1}) \) for the power series to invoke Yurinskii’s coupling in order to establish the uniform inference of the series estimator, where \( a_n \) is a slowly divergent positive sequence. It is clear that our restrictions on \( m_n \) are almost the same as those in Belloni, Chernozhukov, Chetverikov, and Kato (2015).
Lemma B1. Under Assumption B2, Assumption 2(iii) holds with
\[ \delta_{2,n} = \zeta_{4/n}^{1-1/k}m_n^{-1/4} + \zeta_{4/n}^{1/3}m_n^{5/6}n^{-1/6}. \]

Proof of Lemma B1. We use Theorem 1 to prove this lemma. For this purpose, it is sufficient to verify Assumption 1 with \( X_{n,t} = k_n^{-1/2}u_tP(X_t) \) and \( k_n = n \). By Assumption B2(ii), \( V_{n,t} = n^{-1}\sigma_n^2(X_t)P(X_t)P(X_t)\top \). Therefore,
\[ k_n\mathbb{E}[V_{n,t}] = \mathbb{E}\left[ \sigma_n^2(X_t)P(X_t)P(X_t)\top \right] \]
which together with Assumptions B2(ii, iii, v) implies that Assumption 1(i) holds. Let \( v_{n,t} = \sigma_n^2(X_t)P(X_t)P(X_t)\top \). Then by Assumptions B2(iii, v, vi)
\[ \mathbb{E}[|v_{n,t}^{(j,l)}|^\kappa] \leq K\mathbb{E}[|p_j(X_t)p_l(X_t)|^\kappa] \leq K\zeta_n^{2\kappa-2} \]
which implies that \( |v_{n,t}^{(j,l)}|^\kappa \leq K\zeta_n^{2\kappa-2} \) for any \( t \), and any \( j, l \leq m_n \). Therefore by Assumption B2(i), we can use Proposition B1 to deduce that Assumption 1(ii) holds with \( r_n = \zeta_n^{2\kappa-2}m_n^{-1/2} \).

By definition
\[ \sum_{t=1}^{k_n} \mathbb{E}\left[ ||X_{n,t}||^3 \right] = n^{-3/2} \sum_{t=1}^{n} \mathbb{E}\left[ ||u_tP(X_t)||^3 \right] \]
which together with Assumptions B2(ii, iii, iv, v, vi) implies that \( \sum_{t=1}^{k_n} \mathbb{E}\left[ ||X_{n,t}||^3 \right] \leq \zeta_n^{3/2}m_n^{-1/2} \).

Since Assumption 1 holds, by Theorem 1 we verify Assumption 2(iii) with \( \delta_{2,n} = \zeta_n^{1-1/k}m_n^{-1/4} + \zeta_n^{1/3}m_n^{5/6}n^{-1/6} \), which finishes the proof.

Q.E.D.

Lemma B2. Under Assumption B2, Assumption 2(iv) holds with \( \delta_{3,n} = \zeta_n^{2-2/k}m_n^{-1/2} \).

Proof of Lemma B2. Denote \( \eta_t(j,k) \equiv p_j(X_t)p_k(X_t) - \mathbb{E}[p_j(X_t)p_k(X_t)] \). By definition,
\[
\mathbb{E}\left[ ||\hat{Q}_n - Q_n||^2 \right] = \mathbb{E}\left[ n^{-1} \sum_{t=1}^{n} \left( P(X_t)P(X_t)\top - \mathbb{E}\left[ P(X_t)P(X_t)\top \right] \right)^2 \right]
= \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \mathbb{E}\left[ \left( n^{-1} \sum_{t=1}^{n} \eta_t(j,k) \right)^2 \right]
= n^{-2} \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \sum_{t=1}^{n} \mathbb{E}\left[ \eta_t^2(j,k) \right]
+ 2n^{-2} \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathbb{E}\left[ \eta_t(j,k)\eta_s(j,k) \right].
\]

(B.10)

By Assumptions B2(v, vi),
\[
n^{-2} \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \sum_{t=1}^{n} \mathbb{E}\left[ \eta_t^2(j,k) \right] \leq n^{-2} \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \sum_{t=1}^{n} \mathbb{E}\left[ p_j(X_t)^2p_k(X_t)^2 \right] \leq \zeta_n^{2}m_n^{2}\n^{-1}.
\]

(B.11)
Since $(X_t)$ is strong mixing by Assumption B2(i), $(p_j(X_t)p_k(X_t))$ is also strong mixing with the same mixing coefficient $(\varphi_s)_{s=0}^\infty$ for any $(j,k)$. Therefore, by the covariance inequality of the strong mixing process (see, e.g., Corollary 14.3 of Davidson (1994)) and Assumptions B2(i, v, vi),

$$\|E[\eta_t(j,k)\eta_s(j,k)]\| \leq K\varphi_{t-s}^{1-2/\kappa} \|p_j(X_t)p_k(X_t)\|_\kappa^2 \leq K\varphi_{t-s}^{1-2/\kappa} \zeta_n^{-4/\kappa}. \quad (B.12)$$

By (B.12) and the summability condition of the mixing coefficients in Assumption B2(i),

$$n^{-2} \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \sum_{t=2}^{n} \sum_{s=1}^{t-1} |E[\eta_t(j,k)\eta_s(j,k)]| \leq Km_n^2 \zeta_n^{-4/\kappa} n^{-2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \varphi_{t-s}^{1-2/\kappa} = O(\zeta_n^{-4/\kappa} m_n^2 n^{-1}). \quad (B.13)$$

From (B.10), (B.11) and (B.13), we deduce $E[\|\hat{Q}_n - Q_n\|^2] = O(\zeta_n^{-4/\kappa} m_n^2 n^{-1})$, which readily implies the assertion of Lemma B2. 

**Lemma B3.** Under Assumption B2, Assumption 2(v) holds with $\delta_{1,n} = \zeta_n^{-2/\kappa} m_n n^{-1/2}$.

**Proof of Lemma B3.** *Step 1.* By the definition of $\hat{A}_n$, we can decompose the estimation error of $\hat{A}_n$ into three components:

$$\hat{A}_n - A_n = n^{-1} \sum_{t=1}^{n} (\hat{u}_t^2 - u_t^2)P(X_t)P(X_t)^\top$$

$$+ n^{-1} \sum_{t=1}^{n} (u_t^2 - \sigma_n^2(X_t))P(X_t)P(X_t)^\top$$

$$+ n^{-1} \sum_{t=1}^{n} \left( \sigma_n^2(X_t)P(X_t)P(X_t)^\top - E[\sigma_n^2(X_t)P(X_t)P(X_t)^\top] \right)$$

which together with the triangle inequality, Assumption B2(ix), (B.14), (B.28) and (B.29) below implies the assertion of Lemma B3.

*Step 2.* In this step, we show that

$$n^{-1} \sum_{t=1}^{n} (\hat{u}_t^2 - u_t^2)P(X_t)P(X_t)^\top = O_p(\zeta_n^2 m_n^2 n^{-1}) \quad (B.14)$$

under the spectral norm. Since $\hat{u}_t - u_t = h(X_t) - \hat{h}_n(X_t)$,

$$n^{-1} \sum_{t=1}^{n} (\hat{u}_t^2 - u_t^2)P(X_t)P(X_t)^\top = n^{-1} \sum_{t=1}^{n} \left( \hat{h}_n(X_t) - h(X_t) \right)^2 P(X_t)P(X_t)^\top$$

$$- 2n^{-1} \sum_{t=1}^{n} \left( \hat{h}_n(X_t) - h(X_t) \right) u_t P(X_t)P(X_t)^\top. \quad (B.15)$$

34
We next show that the two terms in the right handside of the above equation are $O_p(\zeta_n^2 m_n^{1-2\rho_h} + \zeta_n^2 m_n^2 n^{-1})$ and $O_p(\zeta_n^2 m_n^{3/2-\rho_h} n^{-1/2} + \zeta_n^2 m_n^2 n^{-1})$ under the spectral norm respectively, which together with the triangle inequality and Assumption B2(ix) proves (B.14). By Assumption B2(vi),

$$n^{-1} \sum_{t=1}^{n} \left( \hat{h}_n(X_t) - h(X_t) \right)^2 P(X_t)^\top P(X_t) \leq \zeta_n^2 m_n^{-1} \sum_{t=1}^{n} \left( \hat{h}_n(X_t) - h_0(X_t) \right)^2. \quad (B.16)$$

By Lemma B2 and Assumption B2(ix), $\|\hat{Q}_n - Q_n\| = o_p(1)$. Hence,

$$\lambda_{\min}^{-1}(\hat{Q}_n) + \lambda_{\max}(\hat{Q}_n) \leq K, \quad \text{with probability approaching one.} \quad (B.17)$$

Define $P_n, U_n, H_n$ and $H_n^*$ as in the proof of Theorem 2. Like (A.24) and (A.26), we can show that

$$\hat{b}_n - b_n^* = (P_n^\top P_n)^{-1} \left( P_n^\top U_n \right) + (P_n^\top P_n)^{-1} P_n^\top (H_n - H_n^*), \quad (B.18)$$

and $n^{-1/2} P_n^\top U_n = O_p(m_n^{1/2})$. Then, by (B.17),

$$\left\| (P_n^\top P_n)^{-1} \left( P_n^\top U_n \right) \right\| \leq \lambda_{\min}^{-1}(\hat{Q}_n) \left\| n^{-1} P_n^\top U_n \right\| = O_p(m_n^{1/2} n^{-1/2}). \quad (B.19)$$

By Assumption B2(v) and (B.17),

$$\left\| (P_n^\top P_n)^{-1} P_n^\top (H_n - H_n^*) \right\|^2 \leq \lambda_{\min}^{-1}(\hat{Q}_n) n^{-1} \left\| H_n - H_n^* \right\|^2 = O_p(m_n^{-2\rho_h}). \quad (B.20)$$

By (B.18), (B.19) and (B.20),

$$\left\| \hat{b}_n - b_n^* \right\| = O_p(m_n^{1/2} n^{-1/2} + m_n^{-\rho_h}). \quad (B.21)$$

By Assumption B2(v), (B.17) and (B.21),

$$n^{-1} \sum_{t=1}^{n} \left( \hat{h}_n(X_t) - h(X_t) \right)^2 \leq 2n^{-1} \sum_{t=1}^{n} \left( \hat{h}_n(X_t) - P(X_t)^\top b_n^* \right)^2 + 2n^{-1} \sum_{t=1}^{n} \left( P(X_t)^\top b_n^* - h(X_t) \right)^2 \quad (B.22)$$

$$\leq 2 \lambda_{\max}(\hat{Q}_n) \left\| \hat{b}_n - b_n^* \right\|^2 + 2 \sup_{x \in X} \left| P(x)^\top b_n^* - h(x) \right|^2$$

$$= O_p(m_n^{-2\rho_h} + m_n n^{-1}).$$

Combined with (B.16), this estimate further implies that

$$n^{-1} \sum_{t=1}^{n} \left( \hat{h}_n(X_t) - h(X_t) \right)^2 P(X_t)^\top P(X_t) = O_p(\zeta_n^2 m_n^{1-2\rho_h} + \zeta_n^2 m_n^2 n^{-1}). \quad (B.23)$$
By Assumptions B2(ii, iii, v, vi, vii),

\[
\mathbb{E} \left[ \left\| n^{-1} \sum_{t=1}^{n} (h(X_t) - P(X_t) b_n^*) u_t P(X_t) P(X_t)^\top \right\|^2 \right]
\]

\[
= \sum_{j=1}^{m_n} \sum_{l=1}^{m_n} n^{-2} \sum_{t=1}^{n} \mathbb{E} \left[ u_t^2 p_j^2(X_t) p_l^2(X_t) \right] \leq K \zeta_n^2 m_n^{-2} \rho_n n^{-1}
\]

which together with the Markov inequality implies that

\[
n^{-1} \sum_{t=1}^{n} (h(X_t) - P(X_t) b_n^*) u_t P(X_t) P(X_t)^\top = O_p(\zeta_n m_n^{-1} \rho_n n^{-1/2}) \tag{B.24}
\]

under the Frobenius norm. By Assumptions B2(ii, iii, v, vi),

\[
\sum_{j=1}^{m_n} \sum_{l=1}^{m_n} \mathbb{E} \left[ \left\| n^{-1} \sum_{t=1}^{n} P(X_t) u_t p_j(X_t) p_l(X_t) \right\|^2 \right]
\]

\[
= \sum_{j=1}^{m_n} \sum_{l=1}^{m_n} n^{-2} \sum_{t=1}^{n} \mathbb{E} \left[ u_t^2 p_k^2(X_t) p_j^2(X_t) p_l^2(X_t) \right] \leq K \zeta_n^4 m_n^3 n^{-1}
\]

which together with the Markov inequality implies that

\[
\sum_{j=1}^{m_n} \sum_{l=1}^{m_n} \left\| n^{-1} \sum_{t=1}^{n} P(X_t) u_t p_j(X_t) p_l(X_t) \right\|^2 = O_p(\zeta_n^4 m_n^3 n^{-1}) \tag{B.25}
\]

By (B.21), Assumptions B2(v, iv) and the Cauchy-Schwarz inequality

\[
\left\| n^{-1} \sum_{t=1}^{n} (\hat{h}_n(X_t) - P(X_t) b_n^*) u_t P(X_t) P(X_t)^\top \right\|^2
\]

\[
\leq \left\| \hat{b}_n - b_n^* \right\|^2 \sum_{j=1}^{m_n} \sum_{l=1}^{m_n} \left\| n^{-1} \sum_{t=1}^{n} P(X_t) u_t p_j(X_t) p_l(X_t) \right\|^2
\]

\[
= O_p(\zeta_n^4 m_n^3 \rho_n n^{-1} + \zeta_n^4 m_n^4 n^{-2})
\]

which together with the Markov inequality implies that

\[
n^{-1} \sum_{t=1}^{n} (\hat{h}_n(X_t) - P(X_t) b_n^*) u_t P(X_t) P(X_t)^\top = O_p(\zeta_n^2 m_n^{3/2} \rho_n n^{-1/2} + \zeta_n^2 m_n^2 n^{-1}) \tag{B.26}
\]

under the Frobenius norm. Collecting the results in (B.24) and (B.26), we obtain

\[
n^{-1} \sum_{t=1}^{n} (\hat{h}_n(X_t) - h(X_t)) u_t P(X_t) P(X_t)^\top = O_p(\zeta_n^2 m_n^{3/2} \rho_n n^{-1/2} + \zeta_n^2 m_n^2 n^{-1}) \tag{B.27}
\]
under the Frobenius norm. The assertion in (B.14) follows from (B.15), (B.23), (B.27) and the triangle inequality.

Step 3. In this step we show that

\[
 n^{-1} \sum_{t=1}^{n} (u_t^2 - \sigma_n^2(X_t))P(X_t)P(X_t) = O_p(\zeta_n m_n n^{-1/2}) \tag{B.28}
\]

under the Frobenius norm. By Assumptions B2(i, ii, vi), we get

\[
 n^{-1} \sum_{t=1}^{n} (u_t^2 - \sigma_n^2(X_t))P(X_t)P(X_t) = O_p(\zeta_n m_n n^{-1/2})
\]

which together with the Markov inequality proves (B.28).

Step 4. Let \( \tilde{P}_n(X_t) = \sigma_n(X_t)P(X_t) \). Then under Assumption B2, we can use the same arguments in proving Lemma B2 (but replacing \( P(X_t) \) by \( \tilde{P}_n(X_t) \)) to show that

\[
 n^{-1} \sum_{t=1}^{n} \left( \sigma_n^2(X_t)P(X_t)P(X_t) - E[\sigma_n^2(X_t)P(X_t)P(X_t)] \right) = O_p(\zeta_n m_n n^{-1/2}) \tag{B.29}
\]

under the Frobenius norm. \( Q.E.D. \)

S.B.4.2 The case with serially correlated residuals

The martingale difference assumption on \( u_t \) makes the verification of Assumption 2 relatively easy since the HAC estimation is not needed in this case. In this subsection, we verify the sufficient conditions of Theorem 2 when \( (u_t)_t \) has serial correlation. The following conditions are needed.

Assumption B3. (i) \((X_t^\top, u_t)_t\) is a strictly stationary strong mixing process with mixing coefficient \((\varphi_s)_{s=0}^\infty\) satisfying \( \sum_{s=1}^\infty s^{r_2} \varphi_s^{(\kappa-4)/(5\kappa)} \) for some finite constants \( \kappa > 5 \) and \( r_2 > 0 \); (ii) the eigenvalues of \( Q_n \) and \( A_n \) are bounded from above and away from zero; (iii) \( \mathbb{E} [\|u_t\|^k ] \leq C < \infty \) almost surely for any \( t \); (iv) \( \max_{1 \leq k \leq m_n} \sup_{x \in \mathcal{X}} |p_k(x)| \leq \zeta_n \) where \( \zeta_n \) is a non-decreasing positive sequence and \( \log(\zeta_n^L) = O(\log(m_n)) \); (v) there exist \( \rho_n > 0 \) and \( b^*_n \in \mathbb{R}^{m_n} \) such that

\[
 \sup_{x \in \mathcal{X}} |P(x)^\top b^*_n - h(x)| = O(m_n^{-\rho_n})
\]

(iii) \( \inf_{x \in \mathcal{X}} \|P(x)\| \geq c \) for all \( m_n \) and some constant \( c > 0 \); (vii) \( n^{1/2} m_n^{-\rho_n} + (\zeta_n + m_n^{1/2}) M_n \zeta_n m_n^{3/2} n^{-1/2} + \zeta_n^{-1/2} m_n^{5/6} n^{-1/6} = o(1) \) and \( \zeta_n^{2/3} M_n^{-r_1 \wedge r_2} = o(1) \) where \( r_1 \) is defined in Assumption 6.

Assumption B3(i) assumes that both the nonparametric regressor \( X_t \) and the residual \( u_t \) are (weakly) dependent processes. Assumption B3(ii) is the same as Assumption 2(ii), which is a standard regularity condition in the series estimation literature. Assumption B3(iii) imposes moment
bound on the residual \( u_t \). Assumptions B3(i, iii) imply a trade-off between the dependence and the moment of the residual \( u_t \). When the dependence is weak, Assumptions B3(i) holds with smaller \( \kappa \) (such as \( \kappa \) close to 5) and vice versa. Assumptions B3(iv, v, vi) are the same as Assumption B2(vi, vii, viii). Assumption B3(vii) specifies the growth rate of \( m_n \) and the bandwidth \( M_n \) in the HAC estimation.

**Proposition B4.** The sufficient conditions of Theorem 2(a) hold under Assumptions 6 and B3.

**Proof of Proposition B3.** First, \( \log(\zeta_n^L) = O(\log(m_n)) \) is maintained in Assumption B3(iv). By Assumption B3(v, vi),

\[
\sup_{x \in X} \frac{n^{1/2} |h(x) - P(x)^\top b_n^*|}{\|P(x)\|} \leq O(n^{1/2}m_n^{-\rho_h}). \tag{B.30}
\]

Therefore, Assumption 2(i) holds with \( \delta_{1,n} = n^{1/2}m_n^{-\rho_h} \), where \( \delta_{1,n} = o(1) \) under Assumption B3(vii). Assumption 2(ii) is directly assumed in Assumption B3(ii). Assumptions 2(iii), 2(iv) and 2(v) have been verified in Lemma B4, Lemma B5 and Lemma B6, respectively; \( \delta_{j,n} = o(1), j \in \{2, 3, 4\} \), holds because of Assumption B3(vii). Therefore, Assumption 2 holds under Assumption B3. Moreover, in Lemma B2 and Lemma B3, we show that Assumptions 2(iv, v) hold with

\[
\delta_{3,n} = \zeta_n^{2-2/\kappa}m_n^{-1/2} \quad \text{and} \quad \delta_{4,n} = \zeta_n^2m_n(M_n^{-1/2} + M_n^{-r_1 \wedge r_2}) + M_n\zeta_n^{3/2}m_n^{-1/2}.
\]

which combined with Assumption B3(vii) implies that \( m_n^{1/2}(\delta_{3,n} + \delta_{4,n}) = o(1) \) holds. \( Q.E.D. \)

Proposition B4 implies that Theorem 2(a) holds with

\[
\delta_n = n^{1/2}m_n^{-\rho_h} + (\zeta_n + m_n^{1/2})M_n\zeta_n m_n^{3/2}n^{-1/2} + \zeta_n^{1-2/\kappa}m_n^{5/6}n^{-1/6} + \zeta_n^2m_n^{3/2}M_n^{-r_1 \wedge r_2} = o(1).
\]

The uniform size control in Theorem 2(b) further require that \( \delta_n(\log m_n)^{1/2} = o(1) \). We next discuss this restriction for some popular approximating functions when the spectral density of the process \( (u_tP(X_t))_t \) is sufficiently smooth, i.e., \( r_1, r_2 \geq 3 \).

For splines and trigonometric series, \( \zeta_n \leq K \). Therefore, Assumption B2(vii) is implied by

\[
n^{1/2}m_n^{-\rho_h} + M_n\zeta_n m_n^{5/6}n^{-1/6} + m_n^{3/2}M_n^{-r_1 \wedge r_2} = o(1). \tag{B.31}
\]

Letting \( M_n = m_n^\alpha \) for any \( \alpha \in (3/(2(r_1 \wedge r_2)), 1/2] \), condition in (B.31) is reduced to

\[
n^{1/2}m_n^{-\rho_h} + m_n^{5/2}n^{-1/2} + m_n^{5/6}n^{-1/6} = o(1).
\]

The above restriction requires that \( m_n = o(n^{1/5}) \) and \( \rho_h > 5/2 \), which is the same condition obtained in the martingale difference case. Moreover the sufficient conditions of Theorem 2(b)
hold if \( m_n = o(n^{1/5} \log n)^{-1/2} \) and \( \rho_n > 5/2 \). For the Legendre polynomials, \( \zeta_n \leq K m_n^{1/2} \). In this case, Assumption B2(vii) is implied by
\[
n^{1/2} m_n^{-\rho_n} + M_n m_n^{5/2} n^{-1/2} + m_n^{4/3-1/\kappa} n^{-1/6} + m_n^{5/2} M_n^{-r_1 \wedge r_2} = o(1). \tag{B.32}
\]
Letting \( M_n = m_n^\alpha \) for any \( \alpha \in (5/(2(r_1 \wedge r_2)), 5/6] \), condition in (B.32) is reduced to
\[
n^{1/2} m_n^{-\rho_n} + m_n^{7/2} n^{-1/2} + m_n^{4/3-1/\kappa} n^{-1/6} = o(1). \tag{B.33}
\]
The above restriction requires that \( m_n = o(n^{1/8}) \) and \( \rho_n > 4 \) for any \( \kappa > 5 \), which is stronger than the condition obtained in the martingale difference case. On the other hand, when the data dependence is weak such that \( \kappa \) is close to 5, the restriction in (B.33) is implied by \( m_n = o(n^{1/7}) \) and \( \rho_n > 5/3 \).

**Lemma B4.** Under Assumption B3, Assumption 2(iii) holds with \( \delta_{3,n} = \zeta_n^{1-2/\kappa} m_n^{5/6} n^{-1/6} \).

**Proof of Lemma B4.** We use Proposition B2 to prove this Lemma. For this purpose, it is sufficient to verify Assumption B1 with \( \rho_n = n \). By Assumptions B3(ii, iii)
\[
\sup_{t,j} \| \varepsilon_t^{(j)} \| \leq \sup_{t,j} \| u_t p_j(X_t) \| \leq K^{1/\kappa} \sup_t \| p_j(X_t) \| \leq K^{1/\kappa} (\lambda_{\max}(Q_n))^{1/\kappa} \zeta_n^{1-2/\kappa}
\]
which verifies Assumption B1(i) with \( c_{\kappa,n} = K^{1/\kappa} \zeta_n^{1-2/\kappa} \). Assumptions B1(ii), B1(iii) and B1(iv) are implied by Assumptions B3(i), B3(ii) and B3(vii), respectively. The assertion of Lemma B4 then follows from Proposition B2. \( Q.E.D. \)

**Lemma B5.** Under Assumption B3, Assumption 2(iv) holds with \( \delta_{3,n} = \zeta_n^{2-2/\kappa} m_n n^{-1/2} \).

The proof of the above lemma follows the same arguments in the proof of Lemma B2, and hence is omitted.

**Lemma B6.** Under Assumptions 6 and B3, Assumption 2(v) holds with
\[
\delta_{4,n} = \zeta_n^2 m_n (M_n m_n^{-1/2} + M_n^{-r_1 \wedge r_2}) + M_n \zeta_n m_n^{3/2 + n^{-1/2}}.
\]

**Proof of Lemma B6.** Step 1. We use Theorem 6 to prove this lemma. In order to cast the setting into that of Theorem 6, we set \( k_n = n \), \( Z_t = (Y_t, X_t^\top) \), \( \theta_0 = h(\cdot) \), \( \hat{\theta}_n = \hat{h}_n(\cdot) \) and
\[
X_{n,t} = n^{-1/2} P(X_t) u_t, \quad \hat{X}_{n,t} = n^{-1/2} P(X_t) (Y_t - \hat{h}_n(X_t)). \tag{B.34}
\]
In this step, we verify that the \( X_{n,t} \) array satisfies Assumption 7. Under Assumption B3, we can use the same arguments in the proof of Proposition B2 to show that the array \( (X_{n,t}) \) satisfies
Assumption 5 with $\bar{c}_n = 6K^{1/\kappa}\zeta_n$, $q = 5\kappa/(\kappa + 1)$ and $\psi_s = \varphi_s^{(\kappa-4)/(5\kappa)}$. It remains to verify conditions (i)–(iv) in Assumption 7.

By (2.5), $\mathbb{E}[X_{n,t}] = 0$ for any $t$ and any $n$. Moreover, by Assumption B3(i), $\mathbb{E}[X_{n,t}X_{n,t+j}] = n^{-1}\mathbb{E}[u_tu_{t+j}P(X_t)P(X_{t+j})^\top]$ only depends on $n$ and $j$. Therefore, Assumption 7(i) holds. Let $\mathcal{F}_{n,t}$ be the $\sigma$-field generated by $\{X_s, u_{s-1}\}_{s \leq t}$. We can use the same argument in the proof of Theorem 14.2 of Davidson (1994) to deduce that

$$\|\mathbb{E}[X_{n,t}^{(l)}X_{n,t+j}^{(k)} | \mathcal{F}_{n,t}] - \mathbb{E}[X_{n,t}^{(l)}X_{n,t+j}^{(k)}]\|_2 \leq 6\varphi_s^{\kappa/2}\|X_{n,t}^{(l)}X_{n,t+j}^{(k)}\|^{\kappa/2}. \tag{B.35}$$

By the definition of $X_{n,t}^{(l)}$ and $X_{n,t+j}^{(k)}$, and Assumptions B3(iii, iv),

$$\|X_{n,t}^{(l)}X_{n,t+j}^{(k)}\|^{\kappa/2} \leq \|u_tu_{t+j}\|^{\kappa/2} \zeta_n^2 n^{-1} \leq K^{2/\kappa}\zeta_n^2 n^{-1} \leq \bar{c}_n^2 n^{-1}, \tag{B.36}$$

which verifies Assumption 7(iii). Furthermore, this estimate and (B.35) imply that

$$\|\mathbb{E}[X_{n,t}^{(l)}X_{n,t+j}^{(k)} | \mathcal{F}_{n,t}] - \mathbb{E}[X_{n,t}^{(l)}X_{n,t+j}^{(k)}]\|_2 \leq 6K^{2/\kappa}\zeta_n^2\varphi_s^{1/2-2/\kappa}n^{-1} \leq \bar{c}_n^2 n^{-1}\varphi_s^{\kappa-2)/(2\kappa)}. \tag{B.37}$$

Since $(\kappa - 4)/(5\kappa) \leq (\kappa - 2)/(2\kappa)$, this estimate implies Assumption 7(ii) with $\psi_s$ defined as above.

Finally, we verify Assumption 7(iv). Under Assumption B3(i), $\psi_s$ is summable. Hence, there exists a finite $\bar{s}$ such that $\psi_s \leq s^{-1}$ for any $s \geq \bar{s}$; otherwise, we could extract a subsequence from $\psi_s$ that is not summable. Therefore,

$$\sup_{s \geq 0} s\psi_s^2 \leq 1 + \max_{0 \leq s \leq \bar{s}} s\psi_s^2 < \infty.$$

Further note that $\sum_{s=0}^{\infty} s^{r}\psi_s < \infty$ holds by Assumption B3(i). This verifies Assumption 7(iv).

**Step 2.** In this step, we finish the proof of Lemma B6 by verifying Assumption 8 for which we note from (B.34) that the $g(\cdot)$ function is defined implicitly as $g(Z_t, h) = (Y_t - h(X_t))P(X_t)$. Hence, by Assumption B3(iv) and (B.22) (which can be proved by Assumption B3 and similar arguments in the proof of Lemma B3)

$$n^{-1} \sum_{t=1}^{n} \left\|g(Z_n, \hat{h}_n) - g(Z_t, h)\right\|_2^2 = n^{-1} \sum_{t=1}^{n} \left(\hat{h}_n(X_t) - h(X_t)\right)^2 P(X_t)^\top P(X_t) \leq \zeta_n^2 m_n n^{-1} \sum_{t=1}^{n} \left(\hat{h}_n(X_t) - h_0(X_t)\right)^2$$

$$= O_p(\zeta_n^2 m_n^{1-2\rho_b} + \zeta_n^2 m_n^{1}\zeta_n^2 n^{-1}),$$

which verifies Assumption 8(i) with $\delta_{\theta, n} = \zeta_n m_n^{1/2-\rho_b} + \zeta_n m_n n^{-1/2}$. By Assumption B3(ii, iii),

$$\|g(Z_t, h)\|_2^2 = \mathbb{E}\left[u_t^2 P(X_t)^\top P(X_t)\right] \leq K^{2/\kappa} \text{Tr}(Q_n) \leq K m_n,$$

which implies Assumption 8(ii). This finishes the proof. Q.E.D.
References


