Online Supplementary Appendices to "Time-Varying Multivariate Causal Processes"

*Jiti Gao and *Bin Peng and †Wei Biao Wu and ‡Yayi Yan

*Department of Econometrics and Business Statistics, Monash University

†Department of Statistics, The University of Chicago

‡School of Statistics and Management, Shanghai University of Finance and Economics

The file includes Appendix A and Appendix B. We first present some technical tools in Appendix A.1, which will be repeatedly used in the development. We next consider estimation and inference of structural impulse responses for time-varying VARMA models in Appendix A.2. We provide the proofs of main results in Appendix A.3. We provide several preliminary lemmas in Appendix B.1 as well as some secondary lemmas in Appendix B.2, and then present the proofs of preliminary lemmas in Appendix B.3. Appendix B.4 discusses several computational issues of the local linear ML estimation. Appendix B.5 reports some additional simulation results.

In what follows, M and $O(1)$ always stand for some bounded constants, and may be different at each appearance.

Appendix A

A.1 Technical Tools

Projection Operator: Define the projection operator

$$\mathscr{P}_t(\cdot) = E[\cdot \mid \mathscr{F}_t] - E[\cdot \mid \mathscr{F}_{t-1}],$$

where $\mathscr{F}_t = \sigma(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \ldots)$. By the Jensen's inequality and the stationarity of $\widetilde{\mathbf{x}}_t(\tau)$, for $l \geq 0$, we have

$$\begin{aligned}
\|\mathscr{P}_{t-l}(\widetilde{\mathbf{x}}_t(\tau))\|_r &= \|E[\widetilde{\mathbf{x}}_t(\tau) \mid \mathscr{F}_{t-l}] - E[\widetilde{\mathbf{x}}_t(\tau) \mid \mathscr{F}_{t-l-1}]\|_r \\
&= \|E[\widetilde{\mathbf{x}}_t(\tau) \mid \mathscr{F}_{t-l}] - E[\widetilde{\mathbf{x}}_t^{(t-l,*)}(\tau) \mid \mathscr{F}_{t-l-1}]\|_r \\
&= \|E[\widetilde{\mathbf{x}}_t(\tau) - \widetilde{\mathbf{x}}_t^{(t-l,*)}(\tau) \mid \mathscr{F}_{t-l}]\|_r \\
&\leq \|\widetilde{\mathbf{x}}_t(\tau) - \widetilde{\mathbf{x}}_t^{(t-l,*)}(\tau)\|_r = \delta_r^{\mathbf{x}(\tau)}(l),
\end{aligned}$$

where $\widetilde{\mathbf{x}}_t^{(t-l,*)}(\tau)$ is a coupled version of $\widetilde{\mathbf{x}}_t(\tau)$ with ε_{t-l} replaced by ε_{t-l}^*.

The Class $\mathscr{H}(C, \boldsymbol{\chi}, M)$:

Recall that we have defined $\boldsymbol{\Theta}_r$ in Assumption 1. Let $\boldsymbol{\chi} = \{\chi_j\}_{j=1}^\infty$ be a sequence of nonnegative real numbers with $|\boldsymbol{\chi}|_1 := \sum_{j=1}^\infty \chi_j < \infty$ and $M > 0$ be some finite constant. Let $|\mathbf{z}|_{\boldsymbol{\chi}} := \sum_{j=1}^\infty \chi_j |\mathbf{z}_j|$ for any $\mathbf{z} \in (\mathbb{R}^m)^\infty$ and $C \geq 1$, where $\mathbf{z}_j$ is the j^{th} column of $\mathbf{z}$. A function $g(\mathbf{z}, \boldsymbol{\vartheta}) : (\mathbb{R}^m)^\infty \times \boldsymbol{\Theta}_r \to \mathbb{R}$ is in class $\mathscr{H}(C, \boldsymbol{\chi}, M)$ if

$$\begin{aligned}
&\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_r} |g(\mathbf{0}, \boldsymbol{\vartheta})| \leq M, \\
&\sup_{\mathbf{z}} \sup_{\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}'} \frac{|g(\mathbf{z}, \boldsymbol{\vartheta}) - g(\mathbf{z}, \boldsymbol{\vartheta}')|}{|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'|(1 + |\mathbf{z}|_{\boldsymbol{\chi}}^C)} \leq M, \\
&\sup_{\boldsymbol{\vartheta}} \sup_{\mathbf{z} \neq \mathbf{z}'} \frac{|g(\mathbf{z}, \boldsymbol{\vartheta}) - g(\mathbf{z}', \boldsymbol{\vartheta})|}{|\mathbf{z} - \mathbf{z}'|_{\boldsymbol{\chi}}(1 + |\mathbf{z}|_{\boldsymbol{\chi}}^{C-1} + |\mathbf{z}'|_{\boldsymbol{\chi}}^{C-1})} \leq M.
\end{aligned}$$

If g is vector- or matrix-valued, $g \in \mathscr{H}(C, \boldsymbol{\chi}, M)$ means that every component of g is in $\mathscr{H}(C, \boldsymbol{\chi}, M)$.

Analytical Gradient:

Let

$$\ell(\mathbf{x}, \mathbf{z}; \boldsymbol{\vartheta}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta}))^\top \mathbf{M}^{-1}(\mathbf{z}; \boldsymbol{\vartheta})(\mathbf{x} - \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta})) - \frac{1}{2}\log\det(\mathbf{M}(\mathbf{z}; \boldsymbol{\vartheta})),$$

where $\mathbf{M}(\mathbf{z}; \boldsymbol{\vartheta}) = \mathbf{H}(\mathbf{z}; \boldsymbol{\vartheta})\mathbf{H}(\mathbf{z}; \boldsymbol{\vartheta})^\top$. Then the first partial derivative is as follows:

$$\frac{\partial \ell(\mathbf{x}, \mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_i} = (\mathbf{x} - \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta}))^\top \mathbf{M}^{-1}(\mathbf{z}; \boldsymbol{\vartheta}) \frac{\partial \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta})}{\partial \vartheta_i}$$

$$-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))^\top \frac{\partial \mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta})}{\partial \vartheta_i}(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))$$
$$-\frac{1}{2}\mathrm{tr}\left(\mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta})\frac{\partial \mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})}{\partial \vartheta_i}\right), \tag{A.1.1}$$

where ϑ_i is the i^{th} element of $\boldsymbol{\vartheta}$.

By (A.1.1), the second partial derivative of $\ell(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta})$ is given by

$$\begin{aligned}\frac{\partial^2 \ell(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j} =& (\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))^\top\mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta})\frac{\partial^2\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\\ &-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))^\top\frac{\partial^2\mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))\\ &+(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))^\top\left(\frac{\partial\mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta})}{\partial\vartheta_i}\frac{\partial\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta})}{\partial\vartheta_j}+\frac{\partial\mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta})}{\partial\vartheta_j}\frac{\partial\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta})}{\partial\vartheta_i}\right)\\ &-\left(\frac{\partial\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta})}{\partial\vartheta_j}\right)^\top\mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta})\frac{\partial\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta})}{\partial\vartheta_i}\\ &-\frac{1}{2}\mathrm{tr}\left(\mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta})\frac{\partial^2\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right)-\frac{1}{2}\mathrm{tr}\left(\frac{\partial\mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta})}{\partial\vartheta_j}\frac{\partial\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})}{\partial\vartheta_i}\right).\end{aligned} \tag{A.1.2}$$

A.2 Impulse Responses for Time-varying VARMA Models

Let $\boldsymbol{\Omega}(\tau)=\boldsymbol{\omega}(\tau)\boldsymbol{\omega}^\top(\tau)$. Note that (2.6) can be rewritten as

$$\mathbf{x}_t=\mathbf{v}_t+\boldsymbol{\Phi}_{0,t}\boldsymbol{\varepsilon}_t++\boldsymbol{\Phi}_{1,t}\boldsymbol{\varepsilon}_{t-1}+\cdots,$$

where $\boldsymbol{\Phi}_{0,t}=\boldsymbol{\omega}(\tau_t)$, $\boldsymbol{\Phi}_{j,t}=\mathbf{E}\prod_{m=0}^{i-1}\boldsymbol{\Xi}(\tau_{t-m})\mathbf{S}\boldsymbol{\omega}(\tau_{t-j})$, $\mathbf{E}=[\mathbf{I}_m,\mathbf{0}_{m\times m(p+q-1)}]$, $\mathbf{S}=[\mathbf{I}_m,\mathbf{0}_{m\times m(p-1)},\mathbf{I}_m,\mathbf{0}_{m\times m(q-1)}]^\top$, $\mathbf{v}_t=\mathbf{a}(\tau_t)+\sum_{j=1}^\infty\mathbf{E}\prod_{m=0}^{i-1}\boldsymbol{\Xi}(\tau_{t-m})\mathbf{S}\mathbf{a}(\tau_{t-j})$ and

$$\boldsymbol{\Xi}(\tau)=\left[\begin{array}{cccccccc}\mathbf{A}_1(\tau)&\cdots&\mathbf{A}_{p-1}(\tau)&\mathbf{A}_p(\tau)&\mathbf{B}_1(\tau)&\cdots&\mathbf{B}_{q-1}(\tau)&\mathbf{B}_q(\tau)\\ \mathbf{I}_m&\cdots&\mathbf{0}_m&\mathbf{0}_m&\mathbf{0}_m&\cdots&\mathbf{0}_m&\mathbf{0}_m\\ \vdots&\ddots&\vdots&\vdots&\vdots&\ddots&\vdots&\vdots\\ \mathbf{0}_m&\cdots&\mathbf{I}_m&\mathbf{0}_m&\mathbf{0}_m&\cdots&\mathbf{0}_m&\mathbf{0}_m\\ &&&&\mathbf{0}_m&\cdots&\mathbf{0}_m&\mathbf{0}_m\\ &&\mathbf{0}&&\mathbf{I}_m&\cdots&\mathbf{0}_m&\mathbf{0}_m\\ &&&&\vdots&\ddots&\vdots&\vdots\\ &&&&\mathbf{0}_m&\cdots&\mathbf{I}_m&\mathbf{0}_m\end{array}\right].$$

It can be shown that $\boldsymbol{\Phi}_{j,t}=\boldsymbol{\Phi}_j(\tau_t)+O(1/T)$ with $\boldsymbol{\Phi}_j(\tau)=\mathbf{E}\boldsymbol{\Xi}^j(\tau)\mathbf{S}\boldsymbol{\omega}(\tau)$. Hence, the estimator of $\boldsymbol{\Phi}_{j,t}$ is given by $\widehat{\boldsymbol{\Phi}}_j(\tau)=\mathbf{E}\widehat{\boldsymbol{\Xi}}^j(\tau)\mathbf{S}\widehat{\boldsymbol{\omega}}(\tau)$, where $\widehat{\boldsymbol{\Xi}}(\tau)$ is obtained from $\boldsymbol{\Xi}(\tau)$ by replacing the $\mathbf{A}_i(\tau)$ and $\mathbf{B}_i(\tau)$ by estimators $\widehat{\mathbf{A}}_i(\tau)$ and $\widehat{\mathbf{B}}_i(\tau)$.

We next discuss how to estimate $\boldsymbol{\omega}(\tau)$. Note that we cannot infer the elements in $\boldsymbol{\omega}(\tau)$ unless certain identification restrictions are imposed. Here, we study the impulse response subject to both short-run timing and long-run restrictions.

Under the short-run timing restrictions, $\boldsymbol{\omega}(\cdot)$ is a lower-triangular matrix. Thus, $\widehat{\boldsymbol{\omega}}(\tau)$ is chosen as the lower triangular matrix from the Cholesky decomposition of $\widehat{\boldsymbol{\Omega}}(\tau)$, i.e., $\widehat{\boldsymbol{\Omega}}(\tau)=\widehat{\boldsymbol{\omega}}(\tau)\widehat{\boldsymbol{\omega}}^\top(\tau)$. Alternatively, one can impose the conditions on the long–run impacts of the shocks (i.e., $\boldsymbol{\Phi}(\tau)$ defined below). Specifically, define

$$\boldsymbol{\Phi}(\tau):=\sum_{j=0}^\infty\boldsymbol{\Phi}_j(\tau)=\mathbf{A}_\tau^{-1}(1)\mathbf{B}_\tau(1)\boldsymbol{\omega}(\tau),$$

where the last equality follows in an obvious matter. Thus, the elements of $\boldsymbol{\Phi}(\tau)$ may be recovered from $\boldsymbol{\Phi}(\tau)\boldsymbol{\Phi}^\top(\tau)=\mathbf{A}_\tau^{-1}(1)\mathbf{B}_\tau(1)\boldsymbol{\Omega}(\tau)\left[\mathbf{A}_\tau^{-1}(1)\mathbf{B}_\tau(1)\right]^\top$. It is then convenient to assume that $\boldsymbol{\Phi}(\tau)$ is a lower–triangular matrix, so $\widehat{\boldsymbol{\Phi}}(\tau)$ can be obtained from the Cholesky decomposition of $\widehat{\mathbf{A}}_\tau^{-1}(1)\widehat{\mathbf{B}}_\tau(1)\widehat{\boldsymbol{\Omega}}(\tau)\left[\widehat{\mathbf{A}}_\tau^{-1}(1)\widehat{\mathbf{B}}_\tau(1)\right]^\top$. Under the long–run restrictions, $\widehat{\boldsymbol{\omega}}(\tau)=\widehat{\mathbf{B}}_\tau^{-1}(1)\widehat{\mathbf{A}}_\tau(1)\widehat{\boldsymbol{\Phi}}(\tau)$. We then have the following proposition.

Proposition A.1. *Suppose the conditions of Theorem 2.1 hold. For any fixed integer $j \geq 0$ and any $\tau \in (0,1)$,*

$$\sqrt{Th}\text{vec}\left(\widehat{\mathbf{\Phi}}_j(\tau) - \mathbf{\Phi}_j(\tau) - \frac{1}{2}h^2\tilde{c}_2\mathbf{\Phi}_j^{(2)}(\tau)\right) \to_D N\left(0, \mathbf{\Sigma}_{\mathbf{\Phi}_j}(\tau)\right),$$

where the $\mathbf{C}_{j,i}(\tau)$ matrices, involved in $\mathbf{\Sigma}_{\mathbf{\Phi}_j}(\tau) = [\mathbf{C}_{j,1}(\tau), \mathbf{C}_{j,2}(\tau)]\,\mathbf{\Sigma}_{\boldsymbol{\theta}}(\tau)\,[\mathbf{C}_{j,1}(\tau), \mathbf{C}_{j,2}(\tau)]^\top$, are specified below accordingly.

1. *Under the short–run timing restrictions,*

$$\begin{aligned}
&\mathbf{C}_{0,1}(\tau) = \mathbf{0},\\
&\mathbf{C}_{j,1}(\tau) = (\boldsymbol{\omega}^\top(\tau) \otimes \mathbf{I}_m)\left(\sum_{i=0}^{j-1}\mathbf{S}^\top(\mathbf{\Xi}^\top(\tau))^{j-1-i} \otimes \mathbf{E}\mathbf{\Xi}^i(\tau)\mathbf{E}^\top\right)\left[\mathbf{0}_{m^2(p+q)\times m}, \mathbf{I}_{m^2(p+q)}\right],\ j \geq 1,\\
&\mathbf{C}_{j,2}(\tau) = (\mathbf{I}_m \otimes \mathbf{E}\mathbf{\Xi}^j(\tau)\mathbf{S})\mathbf{L}_m^\top\left(\mathbf{L}_m\mathbf{N}_1(\tau)\mathbf{L}_m^\top\right)^{-1},\ j \geq 0,
\end{aligned}$$

in which $\mathbf{N}_1(\tau) = (\mathbf{I}_{m^2} + \mathbf{K}_{m,m})(\boldsymbol{\omega}(\tau) \otimes \mathbf{I}_m)$, the elimination matrix $\mathbf{L}_m$ satisfies that $\text{vech}(\mathbf{F}) = \mathbf{L}_m\text{vec}(\mathbf{F})$ *for any $m \times m$ matrix $\mathbf{F}$, and the commutation matrix $\mathbf{K}_{m,n}$ satisfies* $\mathbf{K}_{m,n}\text{vec}(\mathbf{G}) = \text{vec}(\mathbf{G}^\top)$ *for any $m \times n$ matrix $\mathbf{G}$.*

2. *Under the long–run restrictions,*

$$\begin{aligned}
\mathbf{C}_{0,1}(\tau) &= (\mathbf{I}_m \otimes \mathbf{\Phi}_0(\tau))\left(\mathbf{N}_1^\top(\tau)\mathbf{N}_1(\tau) + \mathbf{N}_2^\top(\tau)\mathbf{N}_2(\tau)\right)^{-1}\mathbf{N}_2^\top(\tau)\mathbf{D}_2(\tau),\\
\mathbf{C}_{j,1}(\tau) &= \left(\boldsymbol{\omega}^\top(\tau) \otimes \mathbf{I}_m\right)\left(\sum_{i=0}^{j-1}\mathbf{S}^\top(\mathbf{\Xi}^\top(\tau))^{j-1-i} \otimes \mathbf{E}\mathbf{\Xi}^i(\tau)\mathbf{E}^\top\right)\left[\mathbf{0}_{m^2(p+q)\times m}, \mathbf{I}_{m^2(p+q)}\right] + (\mathbf{I}_m \otimes \mathbf{E}\mathbf{\Xi}^j(\tau)\mathbf{S})\\
&\quad \times \left(\mathbf{N}_1^\top(\tau)\mathbf{N}_1(\tau) + \mathbf{N}_2^\top(\tau)\mathbf{N}_2(\tau)\right)^{-1}\mathbf{N}_2^\top(\tau)\mathbf{D}_2(\tau)\left[\mathbf{0}_{m^2(p+q)\times m}, \mathbf{I}_{m^2(p+q)}\right],\ j \geq 1,\\
\mathbf{C}_{j,2}(\tau) &= (\mathbf{I}_m \otimes \mathbf{E}\mathbf{\Xi}^j(\tau)\mathbf{S})\mathbf{L}_m^\top\left(\mathbf{N}_1^\top(\tau)\mathbf{N}_1(\tau) + \mathbf{N}_2^\top(\tau)\mathbf{N}_2(\tau)\right)^{-1}\mathbf{N}_1^\top(\tau)\mathbf{D}_1,\ j \geq 0,
\end{aligned}$$

where $\mathbf{N}_2(\tau) = \mathbf{Q}\left[\mathbf{I}_m \otimes \mathbf{A}_\tau^{-1}(1)\mathbf{B}_\tau(1)\right]$,

$$\begin{aligned}
\mathbf{D}_2(\tau) &= \mathbf{Q}\left\{[\mathbf{\Phi}^\top(\tau) \otimes \mathbf{A}_\tau^{-1}(1)]\nabla_{\boldsymbol{\alpha}(\tau)}\mathbf{A}_\tau(1) - [\boldsymbol{\omega}^\top(\tau) \otimes \mathbf{A}_\tau^{-1}(1)]\nabla_{\boldsymbol{\alpha}(\tau)}\mathbf{B}_\tau(1)\right\},\\
\nabla_{\boldsymbol{\alpha}(\tau)}\mathbf{A}_\tau(1) &= [-\mathbf{I}_{m^2}, ..., -\mathbf{I}_{m^2}, \mathbf{0}_{m^2}, ..., \mathbf{0}_{m^2}]\ (m^2 \times m^2(p+q)),\\
\nabla_{\boldsymbol{\alpha}(\tau)}\mathbf{B}_\tau(1) &= [\mathbf{0}_{m^2}, ..., \mathbf{0}_{m^2}, \mathbf{I}_{m^2}, ..., \mathbf{I}_{m^2}]\ (m^2 \times m^2(p+q)),
\end{aligned}$$

the duplication matrix $\mathbf{D}_1$ satisfies $\text{vec}(\mathbf{\Omega}(\tau)) = \mathbf{D}_1\text{vech}(\mathbf{\Omega}(\tau))$*, and $\mathbf{Q}$ is a $m(m-1)/2 \times m^2$ selection matrix of 0 and 1 such that* $\mathbf{Q}\text{vec}(\mathbf{\Phi}(\tau)) = \mathbf{0}$.

It is easy to see that $\widehat{\mathbf{\Xi}}(\tau) \to_P \mathbf{\Xi}(\tau)$ and $\widehat{\boldsymbol{\omega}}(\tau) \to_P \boldsymbol{\omega}(\tau)$ by Theorem 2.1. As a result, $\widehat{\mathbf{\Sigma}}_{\mathbf{\Phi}_j}(\tau) \to_P \mathbf{\Sigma}_{\mathbf{\Phi}_j}(\tau)$, where $\widehat{\mathbf{\Sigma}}_{\mathbf{\Phi}_j}(\tau)$ has a form identical to $\mathbf{\Sigma}_{\mathbf{\Phi}_j}(\tau)$ but replacing $\mathbf{\Xi}(\tau)$ and $\boldsymbol{\omega}(\tau)$ with their estimators, respectively.

A.3 Proofs of the Main Results

Proof of Proposition 2.1.

(1). In order to construct a solution to $\widetilde{\mathbf{x}}_t(\tau)$, we consider for each fixed $p \geq 0$ and $q > 0$, the approximated p-Markov process $\{\widetilde{\mathbf{x}}_{p,q,t}(\tau)\}_{t\geq 0}$ defined by $\widetilde{\mathbf{x}}_{p,q,t}(\tau) = \mathbf{0}$ for $t \leq -q$ and the recurrence equation

$$\widetilde{\mathbf{x}}_{p,q,t}(\tau) = \boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{p,q,t-1}(\tau), \ldots, \widetilde{\mathbf{x}}_{p,q,t-p}(\tau), \mathbf{0}, \ldots; \boldsymbol{\theta}(\tau)\right) + \mathbf{H}\left(\widetilde{\mathbf{x}}_{p,q,t-1}(\tau), \ldots, \widetilde{\mathbf{x}}_{p,q,t-p}(\tau), \mathbf{0}, \ldots; \boldsymbol{\theta}(\tau)\right)\boldsymbol{\varepsilon}_t$$

for $t > -q$. By Assumption 1, we have

$$\begin{aligned}
&\|\widetilde{\mathbf{x}}_{p,q+1,0}(\tau) - \widetilde{\mathbf{x}}_{p,q,0}(\tau)\|_r\\
&\leq \left(\sum_{j=1}^p \alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_0\|_r\sum_{j=1}^p \beta_j(\boldsymbol{\theta}(\tau))\right)\|\widetilde{\mathbf{x}}_{p,q+1,-j}(\tau) - \widetilde{\mathbf{x}}_{p,q,-j}(\tau)\|_r\\
&\leq \rho(\tau)\|\widetilde{\mathbf{x}}_{p,q+1-j,0}(\tau) - \widetilde{\mathbf{x}}_{p,q-j,0}(\tau)\|_r, \qquad \text{(A.3.1)}
\end{aligned}$$

where the last inequality follows from the definition of $\widetilde{\mathbf{x}}_{p,q,-j}(\tau)$ and $\widetilde{\mathbf{x}}_{p,q-j,0}(\tau)$. Note that these two quantities have the same distribution for each triplet of positive integers (p,q,j).

Now, let $u_k = \|\widetilde{\mathbf{x}}_{p,k+1,0}(\tau) - \widetilde{\mathbf{x}}_{p,k,0}(\tau)\|_r$ and $v_t = \max_{k\geq t} u_k$. Using (A.3.1) and the fact that v_t is a nonincreasing sequence, we have $v_t \leq \rho(\tau) v_{t-p}$ for all $t \geq 1$. Then recursively $v_t \leq \rho(\tau)^{-\lfloor -t/p \rfloor} v_{t+p\lfloor -t/p\rfloor}$. Since $v_{t+p\lfloor -t/p\rfloor} \leq u_0$ and $-\lfloor -t/p\rfloor \geq t/p$, we have $u_t \leq v_t \leq \rho(\tau)^{t/p} u_0$, i.e., $\|\widetilde{\mathbf{x}}_{p,n+1,t}(\tau) - \widetilde{\mathbf{x}}_{p,n,t}(\tau)\|_r = O(\rho(\tau)^{n/p})$. Hence, for each p, $\{\widetilde{\mathbf{x}}_{p,n,0}(\tau)\}_{n\in\mathbb{N}}$ is a Cauchy sequence in the $\mathbb{L}^r$ space; it converges to some $\widetilde{\mathbf{x}}_{p,0}(\tau) \in \mathbb{L}^r$ since this space is complete. From its construction, it is clear that $\widetilde{\mathbf{x}}_{p,n,0}(\tau)$ is measurable with respect to the σ-field generated by $\{\boldsymbol{\varepsilon}_t\}_{t\leq 0}$. The $\mathbb{L}^r$-convergence implies that this is also true for $\widetilde{\mathbf{x}}_{p,0}(\tau)$. Hence, there exists some measurable function $\mathbf{J}_p(\cdot)$ such that $\widetilde{\mathbf{x}}_{p,0}(\tau) = \mathbf{J}_p(\tau, \boldsymbol{\varepsilon}_0, \boldsymbol{\varepsilon}_{-1}, ...)$ and shifting the lag $t \in \mathbb{R}$ leads to the equality $\widetilde{\mathbf{x}}_{p,t}(\tau) = \mathbf{J}_p(\tau, \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, ...)$.

Let $\mu_{p,r}(\tau) = \|\widetilde{\mathbf{x}}_{p,t}(\tau)\|_r$ and $\Delta_{p,r}(\tau) = \|\widetilde{\mathbf{x}}_{p+1,t}(\tau) - \widetilde{\mathbf{x}}_{p,t}(\tau)\|_r$, we then have

$$
\begin{aligned}
\mu_{p,r}(\tau) &\leq \|\widetilde{\mathbf{x}}_{p,t}(\tau) - \widetilde{\mathbf{x}}_{0,t}(\tau)\|_r + \mu_{0,r}(\tau) \\
&\leq \left(\sum_{j=1}^{p} \alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_1\|_r \sum_{j=1}^{p} \beta_j(\boldsymbol{\theta}(\tau)) \right) \mu_{p,r}(\tau) + \mu_{0,r}(\tau),
\end{aligned}
$$

where the second inequality follows from Assumption 1.

Recall that we have defined $\rho(\tau) := \sum_{j=1}^{\infty} \alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_t\|_r \sum_{j=1}^{\infty} \beta_j(\boldsymbol{\theta}(\tau))$ in the body of this proposition. As $0 \leq \rho(\tau) < 1$ by Assumption 1, we have

$$
\sup_{p\geq 0} \mu_{p,r}(\tau) \leq (1-\rho(\tau))^{-1} \mu_{0,r}(\tau) < \infty.
$$

Similarly, we have

$$
\begin{aligned}
\Delta_{p,r}(\tau) \leq & \left(\sum_{j=1}^{p} \alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_1\|_r \sum_{j=1}^{p} \beta_j(\boldsymbol{\theta}(\tau)) \right) \Delta_{p,r}(\tau) \\
& + \left(\alpha_{p+1}(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_1\|_r \beta_{p+1}(\boldsymbol{\theta}(\tau))\right) \|\widetilde{\mathbf{x}}_{p+1,t-p-1}(\tau)\|_r .
\end{aligned}
$$

Hence,

$$
\Delta_{p,r}(\tau) \leq \left(\alpha_{p+1}(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_1\|_r \beta_{p+1}(\boldsymbol{\theta}(\tau))\right) (1-\rho(\tau))^{-2} \mu_{0,r}(\tau) \to 0
$$

as $p \to \infty$.

According to the above development, we are readily to conclude that $\widetilde{\mathbf{x}}_{p,t}(\tau) \to \widetilde{\mathbf{x}}_t(\tau)$ as $p \to \infty$ in the $\mathbb{L}^r$ space. As a limit of strictly stationary process in $\mathbb{L}^r$, $\widetilde{\mathbf{x}}_t(\tau)$ is a stationary process and $\sup_{\tau\in[0,1]} \|\widetilde{\mathbf{x}}_t(\tau)\|_r < \infty$, and $\widetilde{\mathbf{x}}_t(\tau) = \mathbf{J}(\tau, \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, ...)$ is the limit in $\mathbb{L}^r$ of $\widetilde{\mathbf{x}}_{p,t}(\tau) = \mathbf{J}_p(\tau, \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, ...)$.

(2). Let $\{\boldsymbol{\varepsilon}_t^*\}$ be an independent copy of $\{\boldsymbol{\varepsilon}_t\}$. Define the process $\{\widetilde{\mathbf{x}}_{p,t}^*(\tau)\}$, in which the difference is that we use $\boldsymbol{\varepsilon}_t$ when $t \neq 0$, and use $\boldsymbol{\varepsilon}_t^*$ when $t = 0$. In addition, define the process $\{\widetilde{\mathbf{x}}_t^*(\tau)\}$ as $\{\widetilde{\mathbf{x}}_t(\tau)\}$, in which again the difference is that we use $\boldsymbol{\varepsilon}_t$ when $t \neq 0$, and use $\boldsymbol{\varepsilon}_t^*$ when $t = 0$. Further define $u_t = \left\|\widetilde{\mathbf{x}}_{p,t}^*(\tau) - \widetilde{\mathbf{x}}_{p,t}(\tau)\right\|_r$.

By construction, $u_t = 0$ for $t < 0$, and $u_0 = \left\|\widetilde{\mathbf{x}}_{p,0}^*(\tau) - \widetilde{\mathbf{x}}_{p,0}(\tau)\right\|_r = O\left(\|\boldsymbol{\varepsilon}_0^* - \boldsymbol{\varepsilon}_0\|_r\right) = O(1)$. For $t > 0$, Assumption 1 gives that

$$
\left\|\widetilde{\mathbf{x}}_{p,t}(\tau) - \widetilde{\mathbf{x}}_{p,t}^*(\tau)\right\|_r \leq \sum_{j=1}^{p} \left(\alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_t\|_r \beta_j(\boldsymbol{\theta}(\tau))\right) \left\|\widetilde{\mathbf{x}}_{p,t-j}(\tau) - \widetilde{\mathbf{x}}_{p,t-j}^*(\tau)\right\|_r . \tag{A.3.2}
$$

Since $\sum_{j=1}^{p} \left(\alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_t\|_r \beta_j(\boldsymbol{\theta}(\tau))\right) \leq \rho(\tau) < 1$, by a recursion argument, we have $u_t \leq u_0$ for all t.

Now, let $v_t = \max_{k\geq t} u_k$. Using (A.3.2) and the fact that v_t is a nonincreasing sequence, we have $v_t \leq \rho(\tau) v_{t-p}$ for all $t \geq 1$. Then recursively $v_t \leq \rho(\tau)^{-\lfloor -t/p\rfloor} v_{t+p\lfloor -t/p\rfloor}$. Since $v_{t+p\lfloor -t/p\rfloor} \leq u_0$ and $-\lfloor -t/p\rfloor \geq t/p$, we have $u_t \leq v_t \leq \rho(\tau)^{t/p} u_0$, i.e., $\left\|\widetilde{\mathbf{x}}_{p,t}(\tau) - \widetilde{\mathbf{x}}_{p,t}^*(\tau)\right\|_r = O(\rho(\tau)^{t/p})$.

The proof of the first result gives

$$
\|\widetilde{\mathbf{x}}_t(\tau) - \widetilde{\mathbf{x}}_{p,t}(\tau)\|_r \leq \sum_{j=p}^{\infty} \Delta_{p,r} \leq \frac{\mu_r(\tau)}{(1-\rho(\tau))^2} \sum_{j=p}^{\infty} \left(\alpha_{p+1}(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_t\|_r \beta_{p+1}(\boldsymbol{\theta}(\tau))\right).
$$

The same bound holds for the quantity $\|\widetilde{\mathbf{x}}_t^*(\tau) - \widetilde{\mathbf{x}}_{p,t}^*(\tau)\|_r$. Thus,

$$
\begin{aligned}
\|\widetilde{\mathbf{x}}_t(\tau) - \widetilde{\mathbf{x}}_t^*(\tau)\|_r &\leq \|\widetilde{\mathbf{x}}_t(\tau) - \widetilde{\mathbf{x}}_{p,t}(\tau)\|_r + \|\widetilde{\mathbf{x}}_{p,t}(\tau) - \widetilde{\mathbf{x}}_{p,t}^*(\tau)\|_r + \|\widetilde{\mathbf{x}}_t^*(\tau) - \widetilde{\mathbf{x}}_{p,t}^*(\tau)\|_r \\
&= O\left(\rho(\tau)^{t/p} + \sum_{j=p+1}^{\infty} \left(\alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_t\|_r \beta_j(\boldsymbol{\theta}(\tau))\right) \right),
\end{aligned}
$$

which completes the proof. □

Proof of Proposition 2.2.

(1). Write

$$\begin{aligned}\|\widetilde{\mathbf{x}}_t(\tau)-\widetilde{\mathbf{x}}_t(\tau')\|_r &\le \|\boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{t-1}(\tau),\ldots;\boldsymbol{\theta}(\tau)\right)-\boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{t-1}(\tau'),\ldots;\boldsymbol{\theta}(\tau')\right)\|_r\\ &\quad+\|\boldsymbol{\varepsilon}_t\|_r\|\mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau),\ldots;\boldsymbol{\theta}(\tau)\right)-\mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau'),\ldots;\boldsymbol{\theta}(\tau')\right)\|_r\\ &\le \sum_{j=1}^{\infty}\left(\alpha_j(\tau)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau)\right)\|\widetilde{\mathbf{x}}_{t-j}(\tau)-\widetilde{\mathbf{x}}_{t-j}(\tau')\|_r\\ &\quad+M|\tau-\tau'|\sum_{j=1}^{\infty}\chi_j\|\widetilde{\mathbf{x}}_{t-j}(\tau')\|_r,\end{aligned}$$

where the second inequality follows from Assumption 1 and Assumption 2. In view of the stationarity of $\widetilde{\mathbf{x}}_t(\tau)$, rearranging the terms in the above inequality yields that

$$\|\widetilde{\mathbf{x}}_t(\tau)-\widetilde{\mathbf{x}}_t(\tau')\|_r\le M(1-\rho(\tau))^{-1}|\tau-\tau'|\sum_{j=1}^{\infty}\chi_j\|\widetilde{\mathbf{x}}_{t-j}(\tau')\|_r=O(|\tau-\tau'|).$$

(2). Write

$$\begin{aligned}\|\mathbf{x}_t-\widetilde{\mathbf{x}}_t(\tau_t)\|_r &\le \|\boldsymbol{\mu}\left(\mathbf{x}_{t-1},\mathbf{x}_{t-2},\ldots;\boldsymbol{\theta}(\tau_t)\right)-\boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{t-1}(\tau_t),\widetilde{\mathbf{x}}_{t-2}(\tau_t),\ldots;\boldsymbol{\theta}(\tau_t)\right)\|_r\\ &\quad+\|\boldsymbol{\varepsilon}_t\|_r\|\mathbf{H}\left(\mathbf{x}_{t-1},\mathbf{x}_{t-2},\ldots;\boldsymbol{\theta}(\tau_t)\right)-\mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau_t),\widetilde{\mathbf{x}}_{t-2}(\tau_t),\ldots;\boldsymbol{\theta}(\tau_t)\right)\|_r\\ &\le\sum_{j=1}^{\infty}\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right)\left\|\mathbf{x}_{t-j}-\widetilde{\mathbf{x}}_{t-j}(\tau_t)\right\|_r\\ &\le\sum_{j=1}^{\infty}\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right)\left\|\mathbf{x}_{t-j}-\widetilde{\mathbf{x}}_{t-j}(\tau_{t-j}\vee 0)\right\|_r\\ &\quad+\sum_{j=1}^{\infty}\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right)\left\|\widetilde{\mathbf{x}}_{t-j}(\tau_{t-j}\vee 0)-\widetilde{\mathbf{x}}_{t-j}(\tau_t)\right\|_r.\end{aligned}$$

As $\left\|\mathbf{x}_{t-j}-\widetilde{\mathbf{x}}_{t-j}(\tau_{t-j}\vee 0)\right\|_r=0$ for $j\ge t$, by the first result of this proposition, we have

$$\begin{aligned}\|\mathbf{x}_t-\widetilde{\mathbf{x}}_t(\tau_t)\|_r &\le\sum_{j=1}^{t-1}\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right)\left\|\mathbf{x}_{t-j}-\widetilde{\mathbf{x}}_{t-j}(\tau_{t-j})\right\|_r\\ &\quad+M\cdot T^{-1}\sum_{j=1}^{\infty}j\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right).\end{aligned}$$

In addition, as $\left\|\mathbf{x}_1-\widetilde{\mathbf{x}}_1(\tau_1)\right\|_r=O(T^{-1})$ and $\sup_{t\ge 2}\sum_{j=1}^{t-1}\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right)<1$, we have

$$\begin{aligned}\|\mathbf{x}_t-\widetilde{\mathbf{x}}_t(\tau_t)\|_r &\le\sum_{j=1}^{t-1}\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right)O(T^{-1})\\ &\quad+M\cdot T^{-1}\sum_{j=1}^{\infty}j\left(\alpha_j(\tau_t)+\|\boldsymbol{\varepsilon}_t\|_r\beta_j(\tau_t)\right)=O(T^{-1}).\end{aligned}$$

The proof is now complete. □

Proof of Theorem 2.1.

(1). First, we introduce a few notations to facilitate the development. Let $\widehat{\boldsymbol{\eta}}(\tau):=[\widehat{\boldsymbol{\theta}}(\tau)^\top,\widehat{\boldsymbol{\theta}}^\star(\tau)^\top]^\top$, $\boldsymbol{\eta}(\tau):=[\boldsymbol{\theta}(\tau)^\top,h\boldsymbol{\theta}^{(1)}(\tau)^\top]^\top$, and $\mathscr{L}_\tau(\boldsymbol{\eta}):=\mathscr{L}_\tau(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)$ for $\boldsymbol{\eta}=[\boldsymbol{\eta}_1^\top,\boldsymbol{\eta}_2^\top]^\top$. Recall that we have defined $\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}$, and let $\boldsymbol{\nabla}_{\boldsymbol{\eta}}$ be defined similarly with respect to the elements of $\boldsymbol{\eta}$.

By the Taylor expansion, we have

$$\widehat{\boldsymbol{\eta}}(\tau)-\boldsymbol{\eta}(\tau)=-(\boldsymbol{\nabla}^2_{\boldsymbol{\eta}}\mathscr{L}_\tau(\overline{\boldsymbol{\eta}}))^{-1}\boldsymbol{\nabla}_{\boldsymbol{\eta}}\mathscr{L}_\tau(\boldsymbol{\eta}(\tau)),$$

with $\overline{\boldsymbol{\eta}}$ between $\widehat{\boldsymbol{\eta}}(\tau)$ and $\boldsymbol{\eta}(\tau)$. By Lemma B.3.4, we have

$$|\nabla_{\boldsymbol{\eta}}\mathscr{L}_\tau(\boldsymbol{\eta}(\tau)) - \nabla_{\boldsymbol{\eta}}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))| = O_P((Th)^{-1}),$$

where $\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) = T^{-1}\sum_{t=1}^T \ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau) + \boldsymbol{\theta}^{(1)}(\tau)(\tau_t - \tau))K_h(\tau_t - \tau)$.

Then we consider $\nabla_{\boldsymbol{\eta}}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))$. Since each element of $\boldsymbol{\theta}(\tau)$ is in $C^3[0,1]$, we have $\boldsymbol{\theta}(\tau_t) = \boldsymbol{\theta}(\tau) + \boldsymbol{\theta}^{(1)}(\tau)(\tau_t - \tau) + \mathbf{r}(\tau_t)$, where $\mathbf{r}(\tau_t) = \frac{1}{2}\boldsymbol{\theta}^{(2)}(\tau)(\tau_t - \tau)^2 + \frac{1}{6}\boldsymbol{\theta}^{(3)}(\overline{\tau})(\tau_t - \tau)^3$ with $\overline{\tau}$ between τ_t and τ. Let $\widehat{\boldsymbol{K}}((\tau_t - \tau)/h) = [K((\tau_t - \tau)/h), (\tau_t - \tau)/hK((\tau_t - \tau)/h)]^\top$. By the Mean Value Theorem, we have

$$\begin{aligned}
&\nabla_{\boldsymbol{\eta}}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) - \frac{1}{Th}\sum_{t=1}^T \widehat{\boldsymbol{K}}((\tau_t - \tau)/h) \otimes \nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t)) \\
&= -\frac{1}{Th}\sum_{t=1}^T \widehat{\boldsymbol{K}}((\tau_t - \tau)/h) \otimes \left[\nabla^2_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t) - u\mathbf{r}(\tau_t))\mathbf{r}(\tau_t)\right]
\end{aligned}$$

with some $u \in [0,1]$. Since $\nabla^2_{\boldsymbol{\vartheta}}\ell$ is in class $\mathscr{H}(3, \boldsymbol{\chi}, M)$ by Lemma B.2, using Lemma B.8 and $|\tau_t - \tau| \le h$ yields

$$\|\nabla^2_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t) - u\mathbf{r}(\tau_t)) - \nabla^2_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau), \widetilde{\boldsymbol{z}}_{t-1}(\tau); \boldsymbol{\theta}(\tau))\|_1 = O(h).$$

The above analyses plus Lemma B.5 reveal that

$$\begin{aligned}
&\nabla_{\boldsymbol{\eta}}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) - \frac{1}{Th}\sum_{t=1}^T \widehat{\boldsymbol{K}}((\tau_t - \tau)/h) \otimes \nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t)) \\
&= -\frac{1}{2}h^2\frac{1}{Th}\sum_{t=1}^T \widehat{\boldsymbol{K}}((\tau_t - \tau)/h) \otimes \left[\nabla^2_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau), \widetilde{\boldsymbol{z}}_{t-1}(\tau); \boldsymbol{\theta}(\tau)) \cdot \boldsymbol{\theta}^{(2)}(\tau)\left(\frac{\tau_t - \tau}{h}\right)^2\right] + O_P(h^3) \\
&= \frac{1}{2}h^2\int_{-\tau/h}^{(1-\tau)/h} K(u)[u^2, u^3]^\top \mathrm{d}u \otimes \left(-\boldsymbol{\Sigma}(\tau)\boldsymbol{\theta}^{(2)}(\tau)\right) + O_P(h^3).
\end{aligned}$$

By Lemmas B.4 and B.5, we have

$$\nabla_{\boldsymbol{\eta}}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) \to_P 0 \quad \text{and} \quad \sup_{\boldsymbol{\eta}}|\nabla^2_{\boldsymbol{\eta}}\mathscr{L}_\tau(\boldsymbol{\eta}) - E(\nabla^2_{\boldsymbol{\eta}}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}))| \to_P 0.$$

Hence, we have $\nabla^2_{\boldsymbol{\eta}}\mathscr{L}_\tau(\overline{\boldsymbol{\eta}}) \to_P \boldsymbol{\Sigma}(\tau)$ and thus for any $\tau \in [h, 1-h]$, as $Th^7 \to 0$, we have

$$\begin{aligned}
&\sqrt{Th}\left(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau) - \frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\theta}^{(2)}(\tau)\right) \\
&= -\boldsymbol{\Sigma}^{-1}(\tau)\frac{1}{\sqrt{Th}}\sum_{t=1}^T K((\tau_t - \tau)/h)\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t)) + o_P(1).
\end{aligned}$$

In addition, by Lemma B.1, we have $E\left(\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\theta}(\tau))\right) = 0$. To prove this theorem, by the Cramer-Wold device, it suffices to show that for any unit vector $\mathbf{d}$,

$$\frac{1}{\sqrt{Th}}\sum_{t=1}^T K((\tau_t - \tau)/h)\mathbf{d}^\top\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t)) \to_D N\left(\mathbf{0}, \widetilde{v}_0\mathbf{d}^\top\boldsymbol{\Omega}(\tau)\mathbf{d}\right).$$

Note that $\{\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t))\}_t$ is a sequence of martingale differences, we prove the asymptotic normality by using the martingale central limit theorem (Hall and Heyde, 1980). We first consider the convergence of conditional variance. Let $w_t(u) = \frac{1}{\sqrt{Th}}K((\tau_t - \tau)/h)\mathbf{d}^\top\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(u), \widetilde{\boldsymbol{z}}_{t-1}(u); \boldsymbol{\theta}(u))$. By Lemma B.8, we have

$$\begin{aligned}
&\sum_{t=1}^T \|w_t(\tau_t)^2 - w_t(\tau)^2\|_1 \\
&\le \frac{1}{Th}\sum_{t=1}^T K((\tau_t - \tau)/h)^2\|\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t)) - \nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau), \widetilde{\boldsymbol{z}}_{t-1}(\tau); \boldsymbol{\theta}(\tau))\|_2 \\
&\quad \times 2\sup_u \|\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(u), \widetilde{\boldsymbol{z}}_{t-1}(u); \boldsymbol{\theta}(u))\|_2 \\
&= O(h) = o(1).
\end{aligned}$$

In addition, by Proposition 2.1, $\{E\left[(\mathbf{d}^\top \boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(u),\widetilde{\boldsymbol{z}}_{t-1}(u);\boldsymbol{\theta}(u)))^2 \mid \mathscr{F}_{t-1}\right]\}_{t=1}^T$ is a sequence of stationary variables and thus we have

$$
\begin{aligned}
&\sum_{t=1}^T E\left(w_t(\tau)^2 \mid \mathscr{F}_{t-1}\right)\\
=\ & \frac{1}{Th}\sum_{t=1}^T K((\tau_t-\tau)/h)^2 E\left[(\mathbf{d}^\top \boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(u),\widetilde{\boldsymbol{z}}_{t-1}(u);\boldsymbol{\theta}(u)))^2 \mid \mathscr{F}_{t-1}\right]\\
\to_P\ & \widetilde{v}_0 \mathbf{d}^\top \boldsymbol{\Omega}(\tau)\mathbf{d}.
\end{aligned}
$$

We next verify the Lindeberg condition. The sum $\sum_{t=1}^T E\left(w_t^2(\tau_t) I(|w_t(\tau_t)|>v)\right)$ is bounded by

$$
ME\left(\sup_\tau |\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau),\widetilde{\boldsymbol{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))|^2 I(\sup_\tau |\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau),\widetilde{\boldsymbol{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))| > \sqrt{Th}v)\right),
$$

which converges to zero since $\|\sup_\tau |\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau),\widetilde{\boldsymbol{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))|\|_2<\infty$ by Lemma B.8.3. The asymptotic normality is then obtained.

The proof of the first result is now complete.

(2). For notation simplicity, we abbreviate $\ell(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta}), \boldsymbol{\mu}(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta}), \mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})$ to $\ell,\boldsymbol{\mu},\mathbf{M}$ in what follows. Note that

$$
\begin{aligned}
\mathrm{d}\ell &= (\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}\mathrm{d}\boldsymbol{\mu} - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \mathrm{d}\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu}) - \frac{1}{2}\mathrm{tr}\{\mathbf{M}^{-1}\mathrm{d}\mathbf{M}\}\\
&= (\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}^\top}\mathrm{d}\boldsymbol{\vartheta} + \frac{1}{2}((\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}\otimes(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1})\frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\mathrm{d}\boldsymbol{\vartheta}\\
&\quad -\frac{1}{2}\mathrm{vec}(\mathbf{M}^{-1})^\top \frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\mathrm{d}\boldsymbol{\vartheta}.
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
\frac{\partial \ell}{\partial \boldsymbol{\vartheta}}\frac{\partial \ell}{\partial \boldsymbol{\vartheta}^\top} &= \frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}}\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}^\top} + \frac{1}{4}\frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}}\mathrm{vec}(\mathbf{M}^{-1})\mathrm{vec}(\mathbf{M}^{-1})^\top \frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\\
&\quad +\frac{1}{4}\frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}}[\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}\otimes \mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}]\frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\\
&\quad +\frac{1}{2}\frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}}[\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}\otimes(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}]\frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\\
&\quad -\frac{1}{2}\frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}}\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})\mathrm{vec}(\mathbf{M}^{-1})^\top \frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\\
&\quad +\frac{1}{2}\frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}}[(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}\otimes \mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}]\frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\\
&\quad -\frac{1}{4}\frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}}[\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})\otimes \mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})]\mathrm{vec}(\mathbf{M}^{-1})^\top \frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\\
&\quad -\frac{1}{2}\frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}}\mathrm{vec}(\mathbf{M}^{-1})(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}^\top}\\
&\quad -\frac{1}{4}\frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}}\mathrm{vec}(\mathbf{M}^{-1})[(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}\otimes(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{M}^{-1}]\frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}. \qquad \text{(A.3.3)}
\end{aligned}
$$

In addition, if $\boldsymbol{\varepsilon}_t$ is normal distributed, we have $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top \otimes \boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top) = 2\mathbf{N}_m + \mathrm{vec}(\mathbf{I}_m)\mathrm{vec}(\mathbf{I}_m)^\top$ and $E(\mathbf{c}\boldsymbol{\varepsilon}_t^\top \otimes \boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top) = \mathbf{0}$, where $\mathbf{c}$ is independent of $\boldsymbol{\varepsilon}_t$, $2\mathbf{N}_m = \mathbf{I}_{m^2} + \mathbf{K}_{mm}$ and $\mathbf{K}_{mm}$ is a commutation matrix. By (A.3.3), if $\boldsymbol{\varepsilon}_t$ is normal distributed, we have

$$
\begin{aligned}
&\boldsymbol{\Omega}(\tau) = E\left(\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau))\cdot \boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau))^\top\right)\\
&= E\left(\frac{\partial \boldsymbol{\mu}(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau))^\top}{\partial \boldsymbol{\vartheta}}\mathbf{M}^{-1}(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau))\frac{\partial \boldsymbol{\mu}(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau))}{\partial \boldsymbol{\vartheta}^\top}\right)\\
&\quad +\frac{1}{2}E\left(\frac{\partial \mathrm{vec}(\mathbf{M}(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau)))^\top}{\partial \boldsymbol{\vartheta}}[\mathbf{M}^{-1}(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau))\otimes \mathbf{M}^{-1}(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau))]\right.\\
&\quad \left.\times \frac{\partial \mathrm{vec}(\mathbf{M}(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta}(\tau)))}{\partial \boldsymbol{\vartheta}^\top}\right).
\end{aligned}
$$

Next, consider the Hessian matrix.

$$
\begin{aligned}
\mathrm{d}^2\ell = & -\mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}} \mathbf{M}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}^\top}\mathrm{d}\boldsymbol{\vartheta} - \mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}} (\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \otimes \mathbf{M}^{-1}) \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}^\top}\mathrm{d}\boldsymbol{\vartheta} \\
& + \mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \mathrm{vec}(\frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}})^\top}{\partial \boldsymbol{\vartheta}} (\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \otimes \mathbf{I}_m) \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}^\top}\mathrm{d}\boldsymbol{\vartheta} \\
& + \frac{1}{2}\mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}} [\mathbf{M}^{-1} \otimes \mathbf{M}^{-1}] \frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\mathrm{d}\boldsymbol{\vartheta} \\
& - \frac{1}{2}\mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \mathrm{vec}(\frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}})^\top}{\partial \boldsymbol{\vartheta}} (\mathrm{vec}(\mathbf{M}^{-1})^\top \otimes \mathbf{I}_d)\mathrm{d}\boldsymbol{\vartheta} \\
& - \frac{1}{2}\mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}} [\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1} \otimes \mathbf{M}^{-1}] \frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\mathrm{d}\boldsymbol{\vartheta} \\
& - \frac{1}{2}\mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}} (\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \otimes \mathbf{M}^{-1}) \frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\mathrm{d}\boldsymbol{\vartheta} \\
& - \frac{1}{2}\mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\vartheta}} (\mathbf{M}^{-1} \otimes \mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})) \frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\mathrm{d}\boldsymbol{\vartheta} \\
& - \frac{1}{2}\mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}} [\mathbf{M}^{-1} \otimes \mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}] \frac{\partial \mathrm{vec}(\mathbf{M})}{\partial \boldsymbol{\vartheta}^\top}\mathrm{d}\boldsymbol{\vartheta} \\
& + \frac{1}{2}\mathrm{d}\boldsymbol{\vartheta}^\top \frac{\partial \mathrm{vec}(\frac{\partial \mathrm{vec}(\mathbf{M})^\top}{\partial \boldsymbol{\vartheta}})^\top}{\partial \boldsymbol{\vartheta}} (\mathrm{vec}(\mathbf{M}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{M}^{-1}) \otimes \mathbf{I}_d)\mathrm{d}\boldsymbol{\vartheta}. \qquad \text{(A.3.4)}
\end{aligned}
$$

By (A.3.4), if $\boldsymbol{\varepsilon_t}$ is normal distributed, we have

$$
\begin{aligned}
\boldsymbol{\Sigma}(\tau) &= E\left(\boldsymbol{\nabla}^2_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau))\right) \\
= & -E\left(\frac{\partial \boldsymbol{\mu}(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau))^\top}{\partial \boldsymbol{\vartheta}} \mathbf{M}^{-1}(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau)) \frac{\partial \boldsymbol{\mu}(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau))}{\partial \boldsymbol{\vartheta}^\top}\right) \\
& -\frac{1}{2}E\left(\frac{\partial \mathrm{vec}(\mathbf{M}(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau)))^\top}{\partial \boldsymbol{\vartheta}} [\mathbf{M}^{-1}(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau)) \otimes \mathbf{M}^{-1}(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau))]\right. \\
& \left. \times \frac{\partial \mathrm{vec}(\mathbf{M}(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\vartheta}(\tau)))}{\partial \boldsymbol{\vartheta}^\top}\right).
\end{aligned}
$$

Then we have $\boldsymbol{\Omega}(\tau) = -\boldsymbol{\Sigma}(\tau)$ if $\boldsymbol{\varepsilon_t}$ is normal distributed. The proof is now complete. □

Proof of Corollary 2.1.

We first consider $k = 1$. Note that since the coefficient functions $\boldsymbol{\theta}(\cdot)$ is Lipschitz continuous, then $\boldsymbol{\theta}(\tau_{T+k}) - \boldsymbol{\theta}(1) = O(1/T)$ for any bounded k. By Assumption 2.1, Theorem 2.1.1 and Assumption 4.2, we have

$$
\begin{aligned}
\widehat{\mathbf{x}}_{T+1|T} - \mathbf{x}_{T+1|T} &= \boldsymbol{\mu}\left(\mathbf{x}_T, ..., \mathbf{x}_1, \mathbf{0}, \dots; \widehat{\boldsymbol{\theta}}(1)\right) - \boldsymbol{\mu}\left(\mathbf{x}_T, \mathbf{x}_{T-1}, \dots; \boldsymbol{\theta}(\tau_{T+1})\right) \\
&\le |\widehat{\boldsymbol{\theta}}(1) - \boldsymbol{\theta}(\tau_{T+1})| \sum_{j=1}^{\infty} \lambda_j |\mathbf{x}_{T+1-j}| + T^{-1}|\widehat{\boldsymbol{\theta}}(1)| \sum_{j=T+1}^{\infty} j\lambda_j |\mathbf{x}_{T+1-j}| \\
&= O_P(h^2 + 1/\sqrt{Th}) + o_P(1/T) = O_P(h^2 + 1/\sqrt{Th}).
\end{aligned}
$$

Similarly, we have $\widehat{\mathbf{M}}_{T+1|T} - \mathbf{M}_{T+1|T} = O_P\left(h^2 + 1/\sqrt{Th}\right)$.

The case of $k > 1$ can be proved in a similar way by using a recursive argument and using Assumption 1.2, Assumption 2.1, Assumption 4.2 and Theorem 2.1.1. □

Proof of Corollary 2.2.

By Lemma B.5 (2) and the proof of Theorem 2.1, we have

$$
\sup_{\tau\in[0,1]} |\widehat{\boldsymbol{\eta}}(\tau) - \boldsymbol{\eta}(\tau)| = O_P((Th)^{-1/2}h^{-1/2}(\log T)^{1/2}).
$$

In addition, applying Lemma B.3 (4), Lemma B.5 (2) and Lemma B.3 (2) to $g = \boldsymbol{\nabla}^2_{\boldsymbol{\vartheta}}\ell$, we have

$$
\sup_{\tau\in[0,1]} |\widehat{\boldsymbol{\Sigma}}(\tau) - \boldsymbol{\Sigma}(\tau)| = O_P((Th)^{-1/2}h^{-1/2}(\log T)^{1/2} + h) = o_P(1)
$$

as $h(\log T)^2 \to 0$ and $\boldsymbol{\nabla}^2_{\boldsymbol{\vartheta}}\ell \in \mathscr{H}(3, \boldsymbol{\chi}, M)$.

For $\widehat{\mathbf{\Omega}}(\tau)$, as $\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell)^\top \in \mathscr{H}(6,\boldsymbol{\chi},M)$, here we use a different argument to prove the result, which leads to weaker moment conditions.

By Lemma B.3.4 and Lemma B.8.2, we have

$$\begin{aligned}&\widehat{\mathbf{\Omega}}(\tau) - (Th)^{-1}\sum_{t=1}^{T} \boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t),\widetilde{\mathbf{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau))\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t),\widetilde{\mathbf{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau))^\top \widehat{K}\left(\frac{\tau_t-\tau}{h}\right)\\ &= O_P((Th)^{-1/2}h^{-1/2}(\log T)^{1/2} + (Th)^{-1}) = o_P(1),\end{aligned}$$

where $\widehat{K}\left(\frac{\tau_t-\tau}{h}\right) = K\left(\frac{\tau_t-\tau}{h}\right) / \left(T^{-1}\sum_{t=1}^T K\left(\frac{\tau_t-\tau}{h}\right)\right)$.

Define $g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\theta}(\tau)) := \boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))^\top$. By Lemma B.4.1, we have $\sup_{\tau\in[0,1]}\delta_{q/2}^{\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta})|}(j) = O(j^{-(3/2+s)})$.

Define $\mathbf{S}_T(\tau) = \sum_{t=1}^T \left[g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\theta}(\tau)) - E(g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\theta}(\tau)))\right]\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)$ and

$$\mathbf{S}_{k,T} = \sum_{t=1}^{k}\left[g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\theta}(\tau)) - E(g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\theta}(\tau)))\right].$$

By partial summation, we have

$$\mathbf{S}_T(\tau) = \sum_{t=1}^{T-1}\left[\widehat{K}\left(\frac{\tau_t-\tau}{h}\right) - \widehat{K}\left(\frac{\tau_{t+1}-\tau}{h}\right)\right]\mathbf{S}_{t,T} + \widehat{K}\left(\frac{1-\tau}{h}\right)\mathbf{S}_{T,T}.$$

Hence, we have $\sup_{\tau\in[0,1]}|\mathbf{S}_T(\tau)| \le M\max_t|\mathbf{S}_{t,T}|$. Note that $\{\mathscr{P}_{t-j}g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\vartheta})\}_t$ forms a sequence of martingale differences. By the Doob's $\mathbb{L}^q$ maximal inequality, Burkholder inequality and the elementary inequality $(\sum_i |a_i|)^q \le \sum_i |a_i|^q$ for $0 < q \le 1$, we obtain that

$$\begin{aligned}\|\max_t|\mathbf{S}_{t,T}|\|_{q/2} &\le \sum_{l=0}^{\infty}\|\max_{t=1,\ldots,T}|\sum_{s=1}^{t}\mathscr{P}_{s-l}g(\widetilde{\mathbf{y}}_s(\tau_s),\boldsymbol{\theta}(\tau))|\|_{q/2}\\ &\le \sum_{l=0}^{\infty}\frac{q/2}{q/2-1}\|\sum_{s=1}^{T}\mathscr{P}_{s-l}g(\widetilde{\mathbf{y}}_s(\tau_s),\boldsymbol{\theta}(\tau))\|_{q/2}\\ &\le \sum_{l=0}^{\infty}\frac{q/2}{(q/2-1)^2}\left[E\left(\sum_{s=1}^{T}(\mathscr{P}_{s-l}g(\widetilde{\mathbf{y}}_s(\tau_s),\boldsymbol{\theta}(\tau)))^2\right)^{q/4}\right]\\ &\le \frac{q/2}{(q/2-1)^2}\sum_{l=0}^{\infty}\left(\sum_{s=1}^{T}\|\mathscr{P}_{s-l}g(\widetilde{\mathbf{y}}_s(\tau_s),\boldsymbol{\theta}(\tau))\|_{q/2}^{q/2}\right)^{2/q}\\ &\le \frac{q/2}{(q/2-1)^2}T^{2/q}\sum_{l=0}^{\infty}\sup_{\tau\in[0,1]}\delta_{q/2}^{\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta})|}(l)\end{aligned}$$

which shows that $\sup_{\tau\in[0,1]}|\frac{1}{Th}\mathbf{S}_T(\tau)| = O_P(T^{2/q-1}h^{-1}) = o_P(1)$. The result then follows directly by Lemma B.3.2. □

Proof of Theorem 2.2.

We prove this theorem by applying Lemma B.10 to the weak Bahadur representation of $\widehat{\boldsymbol{\theta}}(\tau)$ given in Lemma B.7.

By Lemma B.7, we have

$$\begin{aligned}\sup_{\tau\in[h,1-h]}&\left|\mathbf{C}(\widehat{\boldsymbol{\theta}}(\tau)-\boldsymbol{\theta}(\tau)) - \frac{1}{2}h^2\widetilde{c}_2\mathbf{C}\boldsymbol{\theta}^{(2)}(\tau)\right.\\ &\left.-\frac{1}{T}\sum_{t=1}^{T}(-\mathbf{C}\mathbf{\Sigma}^{-1}(\tau))\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t),\widetilde{\mathbf{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))K_h(\tau_t-\tau)\right|\\ &= O_P(\gamma_T + \beta_T h^2 + h^3 + (Th)^{-1}) = o_P((Th\log T)^{-1/2}) \qquad \text{(A.3.5)}\end{aligned}$$

as $Th^7 \log T \to 0$ and $Th^2/(\log T)^4 \to \infty$. In addition, by Lemmas B.8, B.4.2 and B.10, we have

$$\lim_{T\to\infty} \Pr\left(\sqrt{\frac{Th}{\widetilde{v}_0}} \sup_{\tau\in[h,1-h]} \left|\mathbf{\Sigma}_{\mathbf{C}}^{-1/2}(\tau)\frac{1}{T}\sum_{t=1}^{T}(-\mathbf{C}\mathbf{\Sigma}^{-1}(\tau))\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t),\widetilde{\mathbf{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))K_h(\tau_t-\tau)\right| -B(1/h) \le \frac{u}{\sqrt{2\log(1/h)}}\right) = \exp(-2\exp(-u)). \tag{A.3.6}$$

By (A.3.5) and (A.3.6), the proof is complete. □

Proof of Corollary 2.3.

By the proof of Lemma B.7, we have

$$\sup_{\tau\in[0,1]}\left|\widehat{\boldsymbol{\eta}}(\tau)-\boldsymbol{\eta}(\tau)-\frac{1}{2}h^2\begin{bmatrix}\widetilde{c}_{0,h}(\tau) & \widetilde{c}_{1,h}(\tau)\\ \widetilde{c}_{1,h}(\tau) & \widetilde{c}_{2,h}(\tau)\end{bmatrix}^{-1}\begin{bmatrix}\widetilde{c}_{2,h}(\tau)\\ \widetilde{c}_{3,h}(\tau)\end{bmatrix}\otimes\boldsymbol{\theta}^{(2)}(\tau) -\frac{1}{T}\sum_{t=1}^{T}K_h(\tau_t-\tau)\begin{bmatrix}\widetilde{c}_{0,h}(\tau) & \widetilde{c}_{1,h}(\tau)\\ \widetilde{c}_{1,h}(\tau) & \widetilde{c}_{2,h}(\tau)\end{bmatrix}^{-1}\begin{bmatrix}1\\ \frac{\tau_t-\tau}{h}\end{bmatrix}\otimes(-\mathbf{\Sigma}^{-1}(\tau))\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))\right|$$
$$= O_P((Th)^{-1/2}h^{3/2}(\log T)^{1/2}) + O(h^3).$$

Hence, we have

$$\sup_{\tau\in[0,1]}\left|\widehat{\boldsymbol{\theta}}(\tau)-\boldsymbol{\theta}(\tau)-\frac{1}{2}h^2 b_h(\tau)\boldsymbol{\theta}^{(2)}(\tau)-\frac{1}{T}\sum_{t=1}^{T}(-\mathbf{\Sigma}^{-1}(\tau))\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))\omega_{t,h}(\tau)\right|$$
$$= O_P((Th)^{-1/2}h^{3/2}(\log T)^{1/2}) + O(h^3).$$

By Lemma B.9, there exists i.i.d. k-dimensional standard normal variables $\mathbf{v}_1, ..., \mathbf{v}_T$ such that

$$\sup_{\tau\in[0,1]}\left|\frac{1}{Th}\sum_{t=1}^{T}\omega_{t,h}(\tau)(\nabla_{\boldsymbol{\theta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))-\mathbf{\Omega}^{1/2}(\tau_t)\mathbf{v}_t)\right|$$
$$= O_P\left(\frac{T^{\frac{q(s+3)-4}{2q(2s+3)-4}}(\log T)^{\frac{2(s+1)(q+1)}{q(2s+3)-2}}}{Th}\right) = O_P\left(\frac{(\log T)^2(hT^{\frac{qs+2}{q(2s+3)-2}})^{-1/2}}{(Th)^{1/2}(\log T)^{1/2}}\right)$$
$$= O_P\left(\frac{(\log T)^2(hT^{\nu})^{-1/2}}{(Th\log T)^{1/2}}\right)$$

with $\nu = \frac{qs+2}{q(2s+3)-2}$. Since $\mathbf{\Omega}(\tau)$ is Lipschitz continuous and $\{\mathbf{v}_t\}_{t=1}^T$ is a sequence of i.i.d. normal variables, we have

$$\sup_{\tau\in[0,1]}\left|\frac{1}{Th}\sum_{t=1}^{T}\omega_{t,h}(\tau)(\mathbf{\Omega}^{1/2}(\tau)-\mathbf{\Omega}^{1/2}(\tau_t))\mathbf{v}_t\right|$$
$$= O_P\left(\frac{h(\log T)^{1/2}}{(Th)^{1/2}}\right) = O_P\left(\frac{h\log T}{(Th\log T)^{1/2}}\right).$$

Combining the above analyses, we then complete the proof.

□

Proof of Proposition 2.3.

Note that in this case $\ell, \nabla\ell, \nabla^2\ell$ is in class $\mathscr{H}(2,\boldsymbol{\chi},M)$ as $\mathbf{H}(\mathbf{z};\boldsymbol{\theta}) \equiv \mathbf{H}(\mathbf{0};\boldsymbol{\theta})$ by Lemma B.2. Hence, we only need that the innovation process has $4+s$ moments for some $s > 0$ compared to $6+s$ moments needed in Theorem 2.2.

Consider Assumptions 1–2 first. For notation simplicity, we ignore the time-varying intercept, and rewrite model (2.11) as

$$\widetilde{\mathbf{y}}_t(\tau) = \mathbf{\Gamma}(\tau)\widetilde{\mathbf{y}}_{t-1}(\tau) + \widetilde{\mathbf{u}}_t(\tau),$$

where $\widetilde{\mathbf{y}}_t(\tau)=[\widetilde{\boldsymbol{\eta}}_t^\top(\tau),...,\widetilde{\boldsymbol{\eta}}_{t-q+1}^\top(\tau),\widetilde{\mathbf{x}}_t^\top(\tau),...,\widetilde{\mathbf{x}}_{t-p+1}^\top(\tau)]^\top$, $\widetilde{\mathbf{u}}_t(\tau)=[\widetilde{\mathbf{x}}_t^\top(\tau),\mathbf{0}_{m(q-1)\times 1}^\top,\widetilde{\mathbf{x}}_t^\top(\tau),\mathbf{0}_{m(p-1)\times 1}^\top]^\top$ and

$$\boldsymbol{\Psi}(\tau)=\begin{bmatrix} -\mathbf{B}_1(\tau) & \cdots & -\mathbf{B}_{q-1}(\tau) & -\mathbf{B}_q(\tau) & -\mathbf{A}_1(\tau) & \cdots & -\mathbf{A}_{p-1}(\tau) & -\mathbf{A}_p(\tau) \\ \mathbf{I}_m & \cdots & \mathbf{0}_m & \mathbf{0}_m & \mathbf{0}_m & \cdots & \mathbf{0}_m & \mathbf{0}_m \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_m & \cdots & \mathbf{I}_m & \mathbf{0}_m & \mathbf{0}_m & \cdots & \mathbf{0}_m & \mathbf{0}_m \\ & & & & \mathbf{0}_m & \cdots & \mathbf{0}_m & \mathbf{0}_m \\ & & \mathbf{0} & & \mathbf{I}_m & \cdots & \mathbf{0}_m & \mathbf{0}_m \\ & & & & \vdots & \ddots & \vdots & \vdots \\ & & & & \mathbf{0}_m & \cdots & \mathbf{I}_m & \mathbf{0}_m \end{bmatrix}.$$

Let $\mathbf{E}=[\mathbf{I}_m,\mathbf{0}_{m\times(m(p+q-1))}]$ and $\mathbf{S}=[\mathbf{I}_m,\mathbf{0}_{m\times m(q-1)},\mathbf{I}_m,\mathbf{0}_{m\times m(p-1)}]^\top$, we have

$$\widetilde{\boldsymbol{\eta}}_t(\tau)=\widetilde{\mathbf{x}}_t(\tau)+\sum_{j=1}^{\infty}(\mathbf{E}\boldsymbol{\Psi}^j(\tau)\mathbf{S})\widetilde{\mathbf{x}}_{t-j}(\tau)$$

$$\widetilde{\mathbf{x}}_t(\tau)=\sum_{j=1}^{\infty}(-\mathbf{E}\boldsymbol{\Psi}^j(\tau)\mathbf{S})\widetilde{\mathbf{x}}_{t-j}(\tau)+\widetilde{\boldsymbol{\eta}}_t(\tau)$$

and thus $\boldsymbol{\Gamma}_j(\tau)=-\mathbf{E}\boldsymbol{\Psi}^j(\tau)\mathbf{S}$. Then Assumption 1 is automatically met if $\sum_{j=1}^\infty|\boldsymbol{\Gamma}_j(\tau)|<1$. By using the property of block matrix determinants and $\det(\mathbf{B}_\tau(L))\neq 0$ for all $|L|\leq 1$ (this implies the maximum eigenvalue of left upper $mq\times mq$ matrix in $\boldsymbol{\Gamma}(\tau)$ is less than 1), it can be shown that the maximum eigenvalue of $\boldsymbol{\Gamma}(\tau)$, denoted by ρ, is less than 1 uniformly over $\tau\in[0,1]$. Hence, we have $\alpha_j(\boldsymbol{\theta}(\tau))=|\boldsymbol{\Gamma}_j(\tau)|=O(\rho^j)$ and $\beta_j(\boldsymbol{\theta}(\tau))=0$. In addition, $|\boldsymbol{\Psi}^j-\boldsymbol{\Psi}'^j|=|\sum_{i=1}^{j-1}\boldsymbol{\Psi}^i(\boldsymbol{\Psi}-\boldsymbol{\Psi}')\boldsymbol{\Psi}'^{j-1-i}|=|\boldsymbol{\Psi}-\boldsymbol{\Psi}'|O(\rho^{j-1})$. Then Assumption 2 is met.

However, by using techniques which are more specific to the VARMA models, the condition $\sum_{j=1}^\infty|\boldsymbol{\Gamma}_j(\tau)|<1$ can be weakened to $\det(\mathbf{A}_\tau(L))\neq 0$ for all $|L|\leq 1$. Similar to the above analysis, we have $\widetilde{\mathbf{x}}_t(\tau)=\sum_{j=0}^\infty\boldsymbol{\Phi}_j(\tau)\widetilde{\boldsymbol{\eta}}_{t-j}(\tau)$ with $|\boldsymbol{\Phi}_j(\tau)|=O(\rho^j)$ as $\det(\mathbf{A}_\tau(L))\neq 0$ for all $|L|\leq 1$, which implies that $\|\widetilde{\mathbf{x}}_t(\tau)\|_r<\infty$ and $\delta_r^{\widetilde{\mathbf{x}}(\tau)}(k)=O(\rho^k)$.

For the identification conditions stated in Assumption 3, it is well known that the final form or echelon form is enough to ensure the uniqueness of the VARMA representation.

For verifying Assumption 4, one need the derivatives of $\boldsymbol{\Gamma}_j$. Define $\boldsymbol{\alpha}=-\text{vec}(\mathbf{B}_1,...,\mathbf{B}_q,\mathbf{A}_1,...,\mathbf{A}_p)$. Note that $\text{dvec}(\boldsymbol{\Psi})=(\mathbf{I}_{m(p+q)}\otimes\mathbf{E}^\top)\text{d}\boldsymbol{\alpha}$ and $\text{dvec}(\boldsymbol{\Psi}^j)=(\boldsymbol{\Psi}^\top\otimes\mathbf{I}_{m(p+q)})\text{dvec}(\boldsymbol{\Psi}^{j-1})+(\mathbf{I}_{m(p+q)}\otimes\boldsymbol{\Psi}^{j-1}\mathbf{E}^\top)\text{d}\boldsymbol{\alpha}$, it is easy to show that

$$\frac{\partial\text{vec}(\boldsymbol{\Gamma}_j)}{\partial\boldsymbol{\alpha}^\top}=-\sum_{i=0}^{j-1}\mathbf{S}^\top(\boldsymbol{\Psi}^\top)^{j-1-i}\otimes\mathbf{E}\boldsymbol{\Psi}^j(\tau)\mathbf{E}^\top.$$

Hence, we have $|\frac{\partial\text{vec}(\boldsymbol{\Gamma}_j)}{\partial\boldsymbol{\alpha}^\top}|=O(\rho^{j-1})$ and $|\frac{\partial\text{vec}(\boldsymbol{\Gamma}_j)}{\partial\boldsymbol{\alpha}^\top}-\frac{\partial\text{vec}(\boldsymbol{\Gamma}'_j)}{\partial\boldsymbol{\alpha}'^\top}|=|\boldsymbol{\Psi}-\boldsymbol{\Psi}'|O(\rho^{j-2})$. Similarly, we can verify the conditions imposed on second order derivatives.

The proof is now complete.

□

Proof of Proposition A.1.

Let $\mathbf{A}(\boldsymbol{\theta})$ be a real, differentiable, $m\times n$ matrix function of real $p\times 1$ vector $\boldsymbol{\theta}$. Define $\nabla_{\boldsymbol{\theta}}\mathbf{A}=\frac{\partial\text{vec}(\mathbf{A})}{\partial\boldsymbol{\theta}^\top}$, and thus $\text{vec}(\text{d}\mathbf{A})=\nabla_{\boldsymbol{\theta}}\mathbf{A}\text{d}\boldsymbol{\theta}$.

Let $\boldsymbol{\alpha}(\tau)=\text{vec}\left(\mathbf{A}_1(\tau),...,\mathbf{A}_p(\tau),\mathbf{B}_1(\tau),...,\mathbf{B}_q(\tau)\right)$, $\boldsymbol{\sigma}(\tau)=\text{vech}\left(\boldsymbol{\Omega}(\tau)\right)$ and $\boldsymbol{\phi}(\tau)=[\boldsymbol{\alpha}^\top(\tau),\boldsymbol{\sigma}^\top(\tau)]^\top$. Given the joint distribution of $\boldsymbol{\alpha}(\tau)$ and $\boldsymbol{\sigma}(\tau)$ in Theorem 2.1, Proposition A.1 can be obtained by using the Delta method. By the first-order approximation of $\text{vec}\left(\widehat{\boldsymbol{\Phi}}_j(\tau)\right)$ around $\text{vec}\left(\boldsymbol{\Phi}_j(\tau)\right)$, we have

$$\sqrt{Th}\text{vec}\left(\widehat{\boldsymbol{\Phi}}_j(\tau)-\boldsymbol{\Phi}_j(\tau)\right)\simeq\nabla_{\boldsymbol{\phi}(\tau)}\boldsymbol{\Phi}_j(\tau)\sqrt{Th}\left(\widehat{\boldsymbol{\phi}}(\tau)-\boldsymbol{\phi}(\tau)\right).$$

To complete the proof, we have to derive an analytic form for the derivative $\nabla_{\boldsymbol{\phi}(\tau)}\boldsymbol{\Phi}_j(\tau)$ under each of the identification restrictions. We have two sets of restrictions: (a) $m(m+1)/2$ restrictions implied by $\boldsymbol{\Omega}(\tau)=\boldsymbol{\omega}(\tau)\boldsymbol{\omega}^\top(\tau)$ and (b) additional $m(m-1)/2$ structural restrictions based on short-run or long-run restrictions.

Consider type (a) restrictions. We begin by considering $\text{d}\boldsymbol{\Omega}(\tau)=\text{d}\boldsymbol{\omega}(\tau)\cdot\boldsymbol{\omega}^\top(\tau)+\boldsymbol{\omega}(\tau)\cdot\text{d}\boldsymbol{\omega}^\top(\tau)$. Let $\mathbf{B}$ and $\mathbf{C}$ be $n\times q$ and $q\times r$ matrices, respectively. By $\text{vec}(\mathbf{ABC})=\left[\mathbf{C}^\top\otimes\mathbf{A}\right]\text{vec}(\mathbf{B})$, $\text{vec}(\mathbf{A}^\top)=\mathbf{K}_{m,n}\text{vec}(\mathbf{A})$ and $\mathbf{K}_{m,q}(\mathbf{A}\otimes\mathbf{C})=(\mathbf{C}\otimes\mathbf{A})\mathbf{K}_{n,r}$, we have $\mathbf{N}_1(\tau)\text{vec}(\text{d}\boldsymbol{\omega}(\tau))=\text{vec}(\text{d}\boldsymbol{\Omega}(\tau))$, where $\mathbf{N}_1(\tau)=(\mathbf{I}_{m^2}+\mathbf{K}_{m,m})(\boldsymbol{\omega}(\tau)\otimes\mathbf{I}_m)$. Let $\mathbf{D}_1$ be the duplication matrix such that $\text{vec}[\boldsymbol{\Omega}(\tau)]=\mathbf{D}_1\text{vech}[\boldsymbol{\Omega}(\tau)]$, which follows that $\mathbf{N}_1(\tau)\text{vec}(\text{d}\boldsymbol{\omega}(\tau))=$

$\mathbf{D}_1 \mathrm{d}\boldsymbol{\sigma}(\tau)$ and

$$\mathbf{N}_1(\tau)\triangledown_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\omega}(\tau) = \mathbf{D}_1. \tag{A.3.7}$$

We then illustrate how to combine equation (A.3.7) with gradient equations from type (b) restrictions in order to compute $\triangledown_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau)$.

In the case of short-run timing restrictions, because types (a) and (b) restrictions do not involve $\boldsymbol{\alpha}$, $\triangledown_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau)$ has the form $[\mathbf{0}, \triangledown_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\omega}(\tau)]$. Let $\mathbf{L}_m$ be the elimination matrix defined by $\mathrm{vech}[\boldsymbol{\omega}(\tau)] = \mathbf{L}_m \mathrm{vec}[\boldsymbol{\omega}(\tau)]$. Because $\boldsymbol{\omega}(\tau)$ is lower triangular subject to short-run restrictions, $\mathbf{L}_m^\top$ is a duplication matrix such that $\mathrm{vec}[\boldsymbol{\omega}(\tau)] = \mathbf{L}_m^\top \mathrm{vech}[\boldsymbol{\omega}(\tau)]$. Write

$$\begin{aligned}
\mathbf{N}_1(\tau)\mathrm{vec}(\mathrm{d}\boldsymbol{\omega}(\tau)) &= \mathbf{D}_1 \mathrm{d}\boldsymbol{\sigma}(\tau),\\
\mathbf{L}_m\mathbf{N}_1(\tau)\mathbf{L}_m^\top\mathrm{vech}(\mathrm{d}\boldsymbol{\omega}(\tau)) &= \mathbf{L}_m\mathbf{D}_1 \mathrm{d}\boldsymbol{\sigma}(\tau) = \mathrm{d}\boldsymbol{\sigma}(\tau),\\
\mathrm{vech}(\mathrm{d}\boldsymbol{\omega}(\tau)) &= \left(\mathbf{L}_m\mathbf{N}_1(\tau)\mathbf{L}_m^\top\right)^{-1} \mathrm{d}\boldsymbol{\sigma}(\tau),\\
\mathrm{vec}(\mathrm{d}\boldsymbol{\omega}(\tau)) &= \mathbf{L}_m^\top\left(\mathbf{L}_m\mathbf{N}_1(\tau)\mathbf{L}_m^\top\right)^{-1} \mathrm{d}\boldsymbol{\sigma}(\tau).
\end{aligned}$$

Hence, $\triangledown_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\omega}(\tau) = \mathbf{L}_m^\top\left(\mathbf{L}_m\mathbf{N}_1(\tau)\mathbf{L}_m^\top\right)^{-1}$. Recall that $\triangledown_{\boldsymbol{\phi}(\tau)}\boldsymbol{\Phi}_j(\tau) = [\triangledown_{\boldsymbol{\alpha}(\tau)}\boldsymbol{\Phi}_j(\tau), \triangledown_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\Phi}_j(\tau)]$. For $\triangledown_{\boldsymbol{\alpha}(\tau)}\boldsymbol{\Phi}_j(\tau)$,

$$\begin{aligned}
\triangledown_{\boldsymbol{\alpha}(\tau)}\boldsymbol{\Phi}_j(\tau) &= \frac{\partial \mathrm{vec}\left[\mathbf{E}\boldsymbol{\Xi}^j(\tau)\mathbf{S}\boldsymbol{\omega}(\tau)\right]}{\partial\boldsymbol{\alpha}^\top(\tau)} = (\boldsymbol{\omega}^\top(\tau)\otimes\mathbf{I}_m)\frac{\partial \mathrm{vec}\left[\mathbf{E}\boldsymbol{\Xi}^j(\tau)\mathbf{S}\right]}{\partial\boldsymbol{\alpha}^\top(\tau)}\\
&= (\boldsymbol{\omega}^\top(\tau)\otimes\mathbf{I}_m)(\mathbf{S}^\top\otimes\mathbf{E})\left(\sum_{i=0}^{j-1}(\boldsymbol{\Xi}^\top(\tau))^{j-1-i}\otimes\boldsymbol{\Xi}^i(\tau)\right)\frac{\partial \mathrm{vec}\left[\boldsymbol{\Xi}(\tau)\right]}{\partial\boldsymbol{\alpha}^\top(\tau)}\\
&= (\boldsymbol{\omega}^\top(\tau)\otimes\mathbf{I}_m)\left(\sum_{i=0}^{j-1}\mathbf{S}^\top(\boldsymbol{\Xi}^\top(\tau))^{j-1-i}\otimes\mathbf{E}\boldsymbol{\Xi}^i(\tau)\mathbf{E}^\top\right).
\end{aligned}$$

For $\triangledown_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\Phi}_j(\tau)$,

$$\begin{aligned}
\triangledown_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\Phi}_j(\tau) &= \frac{\partial \mathrm{vec}\left[\mathbf{E}\boldsymbol{\Xi}^j(\tau)\mathbf{S}\boldsymbol{\omega}(\tau)\right]}{\partial\boldsymbol{\sigma}^\top(\tau)} = (\mathbf{I}_m\otimes\mathbf{E}\boldsymbol{\Xi}^j(\tau)\mathbf{S})\frac{\partial \mathrm{vec}\left[\boldsymbol{\omega}(\tau)\right]}{\partial\boldsymbol{\sigma}^\top(\tau)}\\
&= (\mathbf{I}_m\otimes\mathbf{E}\boldsymbol{\Xi}^j(\tau)\mathbf{S})\mathbf{L}_m^\top\left(\mathbf{L}_m\mathbf{N}_1(\tau)\mathbf{L}_m^\top\right)^{-1}.
\end{aligned}$$

In the case of long-run restrictions, type (b) restrictions involve $\boldsymbol{\alpha}(\tau)$, so that $\triangledown_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau)$ has the form $[\triangledown_{\boldsymbol{\alpha}(\tau)}\boldsymbol{\omega}(\tau), \triangledown_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\omega}(\tau)]$. First, equation (A.3.7) must be extended in the form $\mathbf{N}_1\triangledown_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau) = [\mathbf{0}, \mathbf{D}_1]$. Second, long-run restrictions can be expressed as $\mathbf{Q}\mathrm{vec}\left[\mathbf{A}_\tau^{-1}(1)\mathbf{B}_\tau(1)\boldsymbol{\omega}(\tau)\right] = \mathbf{0}$, where $\mathbf{Q}$ is a $m(m-1)/2\times m^2$ selection matrix of 0 and 1, $\mathbf{A}_\tau(1) = \mathbf{I}_m - \sum_{i=1}^p\mathbf{A}_i(\tau)$ and $\mathbf{B}_\tau(1) = \mathbf{I}_m + \sum_{i=1}^q\mathbf{B}_i(\tau)$. By $\mathrm{d}\mathbf{A}^{-1} = -\mathbf{A}^{-1}\cdot\mathrm{d}\mathbf{A}\cdot\mathbf{A}^{-1}$, we have

$$\begin{aligned}
&\mathbf{Q}\mathrm{vec}\left[\mathbf{A}_\tau^{-1}(1)\mathbf{B}_\tau(1)\boldsymbol{\omega}(\tau)\right] = \mathbf{0},\\
&\mathbf{Q}\mathrm{vec}\left[\mathrm{d}(\mathbf{A}_\tau^{-1}(1))\mathbf{B}_\tau(1)\boldsymbol{\omega}(\tau) + \mathbf{A}_\tau^{-1}(1)\mathrm{d}(\mathbf{B}_\tau(1))\boldsymbol{\omega}(\tau) + \mathbf{A}_\tau^{-1}(1)\mathbf{B}_\tau(1)\mathrm{d}\boldsymbol{\omega}(\tau)\right] = \mathbf{0},\\
&\mathbf{Q}\mathrm{vec}\left[-\mathbf{A}_\tau^{-1}(1)\mathrm{d}(\mathbf{A}_\tau(1))\mathbf{A}_\tau^{-1}(1)\mathbf{B}_\tau(1)\boldsymbol{\omega}(\tau) + \mathbf{A}_\tau^{-1}(1)\mathrm{d}(\mathbf{B}_\tau(1))\boldsymbol{\omega}(\tau) + \mathbf{A}_\tau^{-1}(1)\mathbf{B}_\tau(1)\mathrm{d}\boldsymbol{\omega}(\tau)\right] = \mathbf{0},\\
&\mathbf{Q}\left[\mathbf{I}_m\otimes\mathbf{A}_\tau^{-1}(1)\mathbf{B}_\tau(1)\right]\mathrm{vec}\left[\mathrm{d}\boldsymbol{\omega}(\tau)\right] = \mathbf{Q}[\boldsymbol{\Phi}^\top(\tau)\otimes\mathbf{A}_\tau^{-1}(1)]\mathrm{vec}[\mathrm{d}\mathbf{A}_\tau(1)] - \mathbf{Q}[\boldsymbol{\omega}^\top(\tau)\otimes\mathbf{A}_\tau^{-1}(1)]\mathrm{vec}[\mathrm{d}\mathbf{B}_\tau(1)],\\
&\mathbf{Q}\left[\mathbf{I}_m\otimes\mathbf{A}_\tau^{-1}(1)\mathbf{B}_\tau(1)\right]\mathrm{vec}\left[\mathrm{d}\boldsymbol{\omega}(\tau)\right] = \mathbf{Q}\left\{[\boldsymbol{\Phi}^\top(\tau)\otimes\mathbf{A}_\tau^{-1}(1)]\triangledown_{\boldsymbol{\alpha}(\tau)}\mathbf{A}_\tau(1) - [\boldsymbol{\omega}^\top(\tau)\otimes\mathbf{A}_\tau^{-1}(1)]\triangledown_{\boldsymbol{\alpha}(\tau)}\mathbf{B}_\tau(1)\right\}\mathrm{d}\boldsymbol{\alpha}(\tau),\\
&\mathbf{N}_2(\tau)\triangledown_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau) = [\mathbf{D}_2(\tau), \mathbf{0}],
\end{aligned}$$

where $\mathbf{N}_2(\tau) = \mathbf{Q}\left[\mathbf{I}_m\otimes\mathbf{A}_\tau^{-1}(1)\mathbf{B}_\tau(1)\right]$, $\mathbf{D}_2(\tau) = \mathbf{Q}\left\{[\boldsymbol{\Phi}^\top(\tau)\otimes\mathbf{A}_\tau^{-1}(1)]\triangledown_{\boldsymbol{\alpha}(\tau)}\mathbf{A}_\tau(1) - [\boldsymbol{\omega}^\top(\tau)\otimes\mathbf{A}_\tau^{-1}(1)]\triangledown_{\boldsymbol{\alpha}(\tau)}\mathbf{B}_\tau(1)\right\}$, $\triangledown_{\boldsymbol{\alpha}(\tau)}\mathbf{A}_\tau(1) = [-\mathbf{I}_{m^2}, ..., -\mathbf{I}_{m^2}, \mathbf{0}_{m^2}, ..., \mathbf{0}_{m^2}]$ ($m^2\times m^2(p+q)$) and $\triangledown_{\boldsymbol{\alpha}(\tau)}\mathbf{B}_\tau(1) = [\mathbf{0}_{m^2}, ..., \mathbf{0}_{m^2}, \mathbf{I}_{m^2}, ..., \mathbf{I}_{m^2}]$ ($m^2\times m^2(p+q)$). Hence,

$$\begin{aligned}
\triangledown_{\boldsymbol{\phi}(\tau)}\boldsymbol{\omega}(\tau) &= [\triangledown_{\boldsymbol{\alpha}(\tau)}\boldsymbol{\omega}(\tau), \triangledown_{\boldsymbol{\sigma}(\tau)}\boldsymbol{\omega}(\tau)]\\
&= \left[\left(\mathbf{N}_1^\top(\tau), \mathbf{N}_2^\top(\tau)\right)\begin{pmatrix}\mathbf{N}_1(\tau)\\ \mathbf{N}_2(\tau)\end{pmatrix}\right]^{-1}\left[\mathbf{N}_1^\top(\tau), \mathbf{N}_2^\top(\tau)\right]\begin{bmatrix}\mathbf{0} & \mathbf{D}_1\\ \mathbf{D}_2(\tau) & \mathbf{0}\end{bmatrix}\\
&= \left(\mathbf{N}_1^\top(\tau)\mathbf{N}_1(\tau) + \mathbf{N}_2^\top(\tau)\mathbf{N}_2(\tau)\right)^{-1}\left[\mathbf{N}_2^\top(\tau)\mathbf{D}_2(\tau), \mathbf{N}_1^\top(\tau)\mathbf{D}_1\right].
\end{aligned}$$

For $\triangledown_{\boldsymbol{\alpha}(\tau)}\boldsymbol{\Phi}_j(\tau)$,

$$\triangledown_{\boldsymbol{\alpha}(\tau)}\boldsymbol{\Phi}_j(\tau) = \frac{\partial \mathrm{vec}\left[\mathbf{E}\boldsymbol{\Xi}^j(\tau)\mathbf{S}\boldsymbol{\omega}(\tau)\right]}{\partial\boldsymbol{\alpha}^\top(\tau)}$$

$$= (\boldsymbol{\omega}^\top(\tau) \otimes \mathbf{I}_m) \frac{\partial \text{vec}\left[\mathbf{E}\boldsymbol{\Xi}^j(\tau)\mathbf{S}\right]}{\partial \boldsymbol{\alpha}^\top(\tau)} + (\mathbf{I}_m \otimes \mathbf{E}\boldsymbol{\Xi}^j(\tau)\mathbf{S}) \frac{\partial \text{vec}\left[\boldsymbol{\omega}(\tau)\right]}{\partial \boldsymbol{\alpha}^\top(\tau)}$$
$$= (\boldsymbol{\omega}^\top(\tau) \otimes \mathbf{I}_m) \left(\sum_{i=0}^{j-1} \mathbf{S}^\top (\boldsymbol{\Xi}^\top(\tau))^{j-1-i} \otimes \mathbf{E}\boldsymbol{\Xi}^i(\tau)\mathbf{E}^\top \right)$$
$$+ (\mathbf{I}_m \otimes \mathbf{E}\boldsymbol{\Xi}^j(\tau)\mathbf{S}) \left(\mathbf{N}_1^\top(\tau)\mathbf{N}_1(\tau) + \mathbf{N}_2^\top(\tau)\mathbf{N}_2(\tau)\right)^{-1} \mathbf{N}_2^\top(\tau)\mathbf{D}_2(\tau).$$

For $\nabla_{\boldsymbol{\sigma}(\tau)} \boldsymbol{\Phi}_j(\tau)$,

$$\nabla_{\boldsymbol{\sigma}(\tau)} \boldsymbol{\Phi}_j(\tau) = \frac{\partial \text{vec}\left[\mathbf{E}\boldsymbol{\Xi}^j(\tau)\mathbf{S}\boldsymbol{\omega}(\tau)\right]}{\partial \boldsymbol{\sigma}^\top(\tau)} = (\mathbf{I}_m \otimes \mathbf{E}\boldsymbol{\Xi}^j(\tau)\mathbf{S}) \frac{\partial \text{vec}\left[\boldsymbol{\omega}(\tau)\right]}{\partial \boldsymbol{\sigma}^\top(\tau)}$$
$$= (\mathbf{I}_m \otimes \mathbf{E}\boldsymbol{\Xi}^j(\tau)\mathbf{S})\mathbf{L}_m^\top \left(\mathbf{N}_1^\top(\tau)\mathbf{N}_1(\tau) + \mathbf{N}_2^\top(\tau)\mathbf{N}_2(\tau)\right)^{-1} \mathbf{N}_1^\top(\tau)\mathbf{D}_1.$$

The proof is now completed. □

Proof of Proposition 2.4.

Since $\mathbf{H}(\mathbf{z};\boldsymbol{\theta})$ is a positive and diagonal matrix, we have

$$|\mathbf{H}(\mathbf{z};\boldsymbol{\theta}) - \mathbf{H}(\mathbf{z}';\boldsymbol{\theta})| = |(\mathbf{H}^2(\mathbf{z};\boldsymbol{\theta}) - \mathbf{H}^2(\mathbf{z}';\boldsymbol{\theta})) \cdot (\mathbf{H}(\mathbf{z};\boldsymbol{\theta}) + \mathbf{H}(\mathbf{z}';\boldsymbol{\theta}))^{-1}|$$
$$\leq \sum_{j=1}^{\infty} |\boldsymbol{\Psi}_j(\tau)|^{1/2} \cdot |\mathbf{x}_j - \mathbf{x}_j'|.$$

Then Assumption 1 is automatically met if $\|\widetilde{\boldsymbol{\eta}}_t(\tau)\|_r \sum_{j=1}^\infty |\boldsymbol{\Psi}_j(\tau)|^{1/2} < 1$. In addition, as $|\boldsymbol{\Psi}_j(\tau)|$ converges to zero with exponential rate and $\partial h_{i,t}^{1/2}/\partial\theta_i = \frac{1}{2} h_{i,t}^{-1/2} \partial h_{i,t}/\partial\theta_i$, similar to the proof of Proposition 2.3 we can easily verify Assumptions 2 and 4. For the identification conditions of the GARCH process, we refer readers to Proposition 3.4 of Jeantheau (1998), who proves that assuming the minimal representation is enough for ensuring Assumption 3 holds.

However, by using techniques which are more specific to the GARCH models, the condition

$$\|\widetilde{\boldsymbol{\eta}}_t(\tau)\|_r \sum_{j=1}^{\infty} |\boldsymbol{\Psi}_j(\tau)|^{1/2} < 1$$

can be weaken to $\|\widetilde{\boldsymbol{\eta}}_t(\tau)\|_r^2 \sum_{j=1}^\infty |\boldsymbol{\Psi}_j(\tau)| < 1$. Define $\widetilde{\mathbf{y}}_t(\tau) = \widetilde{\mathbf{x}}_t(\tau) \odot \widetilde{\mathbf{x}}_t(\tau)$ and $\widetilde{\mathbf{V}}_t(\tau) = \text{diag}\left(\widetilde{\boldsymbol{\eta}}_t(\tau) \odot \widetilde{\boldsymbol{\eta}}_t(\tau)\right)$. We first prove the existence of $\|\widetilde{\mathbf{y}}_t(\tau)\|_{r/2}$ (which implies the existence of $\|\widetilde{\mathbf{x}}_t(\tau)\|_r$) as well as its weak dependence property by means of a chaotic expansion. Since $\widetilde{\mathbf{y}}_t(\tau) = \widetilde{\mathbf{V}}_t(\tau)\boldsymbol{\alpha}(\tau) + \sum_{j=1}^\infty \widetilde{\mathbf{V}}_t(\tau)\boldsymbol{\Psi}_j(\tau)\widetilde{\mathbf{y}}_{t-j}(\tau)$, by substitute $\widetilde{\mathbf{y}}_{t-j}(\tau)$ recursively, we have

$$\widetilde{\mathbf{y}}_t(\tau) = \widetilde{\mathbf{V}}_t(\tau) \left\{ \boldsymbol{\alpha}(\tau) + \sum_{k=1}^{\infty} \sum_{j_1,..,j_k=1}^{\infty} \boldsymbol{\Psi}_{j_1}(\tau)\widetilde{\mathbf{V}}_{t-j_1}(\tau) \cdots \boldsymbol{\Psi}_{j_k}(\tau)\widetilde{\mathbf{V}}_{t-j_1-\cdots-j_k}(\tau)\boldsymbol{\alpha}(\tau) \right\}.$$

To prove the boundedness of $\|\widetilde{\mathbf{y}}_t(\tau)\|_{r/2}$, since $\{\widetilde{\mathbf{V}}_t(\tau)\}$ are independent random variables, it suffices to show that

$$\sum_{k=1}^{\infty} \sum_{j_1,..,j_k=1}^{\infty} \left\| \boldsymbol{\Psi}_{j_1}(\tau)\widetilde{\mathbf{V}}_{t-j_1}(\tau) \cdots \boldsymbol{\Psi}_{j_k}(\tau)\widetilde{\mathbf{V}}_{t-j_1-\cdots-j_k}(\tau)\boldsymbol{\alpha}(\tau) \right\|_{r/2} < \infty.$$

By using $\sup_{\tau\in[0,1]} |\boldsymbol{\alpha}(\tau)| < \infty$ and $\|\widetilde{\mathbf{V}}_t(\tau)\|_{r/2} \sum_{j=1}^\infty |\boldsymbol{\Psi}_j(\tau)| < 1$, we have

$$\sum_{k=1}^{\infty} \sum_{j_1,..,j_k=1}^{\infty} \left\| \boldsymbol{\Psi}_{j_1}(\tau)\widetilde{\mathbf{V}}_{t-j_1}(\tau) \cdots \boldsymbol{\Psi}_{j_k}(\tau)\widetilde{\mathbf{V}}_{t-j_1-\cdots-j_k}(\tau)\boldsymbol{\alpha}(\tau) \right\|_{r/2}$$
$$\leq \sup_{\tau\in[0,1]} |\boldsymbol{\alpha}(\tau)| \sum_{k=1}^{\infty} \sum_{j_1,..,j_k=1}^{\infty} |\boldsymbol{\Psi}_{j_1}(\tau)| \cdots |\boldsymbol{\Psi}_{j_k}(\tau)| \|\widetilde{\mathbf{V}}_t(\tau)\|_{r/2}^k$$
$$\leq \sup_{\tau\in[0,1]} |\boldsymbol{\alpha}(\tau)| \sum_{k=1}^{\infty} \left(\|\widetilde{\mathbf{V}}_t(\tau)\|_{r/2} \sum_{j=1}^{\infty} |\boldsymbol{\Psi}_j(\tau)| \right)^k < \infty.$$

Hence, we have $\|\widetilde{\boldsymbol{x}}_t(\tau)\|_r < \infty$. Next, we show that $\delta_r^{\widetilde{\mathbf{x}}(\tau)}(k) = O(\rho^k)$ for some $0 < \rho < 1$. Write

$$\widetilde{\mathbf{y}}_t(\tau) = \text{diag}\left(\boldsymbol{\alpha}(\tau) + \sum_{j=1}^{\infty} \boldsymbol{\Psi}_j(\tau)\widetilde{\mathbf{y}}_{t-j}(\tau)\right)(\widetilde{\boldsymbol{\eta}}_t(\tau) \odot \widetilde{\boldsymbol{\eta}}_t(\tau)).$$

By using the same arguments as in the proof of Proposition 2.1, we have $\delta_r^{\widetilde{\mathbf{y}}(\tau)}(k) = O(\rho^k)$ since $\|\widetilde{\boldsymbol{\eta}}_t(\tau)\|_r^2 \sum_{j=1}^{\infty} |\boldsymbol{\Psi}_j(\tau)| < 1$ and $\beta_j(\boldsymbol{\theta}(\tau)) = |\boldsymbol{\Psi}_j(\tau)| = O(\rho^j)$. Since $|a-b| \leq |a^2 - b^2|^{1/2}$ for $a \geq 0, b \geq 0$ and $\mathbf{H}(\cdot)$ is a positive diagonal matrix, for $t \geq 1$, we have

$$\begin{aligned}
\|\widetilde{\boldsymbol{x}}_t(\tau) - \widetilde{\boldsymbol{x}}_t^*(\tau)\|_r &\leq \sum_{j=1}^{t} |\boldsymbol{\Psi}_j(\tau)|^{1/2} \cdot \|\widetilde{\mathbf{y}}_{t-j}(\tau) - \widetilde{\mathbf{y}}_{t-j}^*(\tau)\|_{r/2}^{1/2} \cdot \|\widetilde{\boldsymbol{\eta}}_t(\tau)\|_r \\
&= \sum_{j=1}^{t} O(\rho^{j/2}) O(\rho^{(t-j)/2}) = O(\rho'^t)
\end{aligned}$$

for some $0 < \rho' < 1$.

The proof is now completed. □

Proof of Proposition 2.5.

Proposition 2.5 can be verified in a similar manner as Propositions 2.3 and 2.4. Note that by the proof of Proposition 2.3, we have $\alpha_j(\boldsymbol{\theta}(\tau)) = O(\rho^j)$ for some $0 < \rho < 1$ and $\|\widetilde{\mathbf{x}}_t(\tau)\|_r < \infty$ provided that $\|\widetilde{\mathbf{v}}_t(\tau)\|_r < \infty$. In addition, by the proof of Proposition 2.3, we have $\|\widetilde{\mathbf{v}}_t(\tau)\|_r < \infty$ and $\beta_j(\boldsymbol{\theta}(\tau)) = O(\rho^j)$ for some $0 < \rho < 1$ since $\|\boldsymbol{\Omega}^{1/2}(\tau)\boldsymbol{\varepsilon}_t\|_r^2 \sum_{j=1}^{\infty} |\boldsymbol{\Psi}_j(\tau)| < 1$.

□

Appendix B

B.1 Preliminary Lemmas

First, we define a few notations for better presentation. First, let $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)^\top$, where $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ are the same generic vectors as in (2.4). Let $\widehat{K}(\cdot)$ be a a kernel function being Lipschitz continuous and bounded on $[-1, 1]$.

For $\tau \in [0, 1]$ and $\boldsymbol{\eta} \in \mathbf{E}_T(r) = \boldsymbol{\Theta}_r \times (h \cdot \boldsymbol{\Theta}^{(1)})$, define

$$G_\tau(\boldsymbol{\eta}) := \frac{1}{Th} \sum_{t=1}^{T} \widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \left[g(\mathbf{y}_t, \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) - E(g(\mathbf{y}_t, \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h))\right], \tag{B.1}$$

where $g(\cdot) \in \mathscr{H}(C, \boldsymbol{\chi}, M)$ and $\mathbf{y}_t = (\mathbf{x}_t, \mathbf{z}_{t-1})$. Let $G_\tau^c(\boldsymbol{\eta})$, $\widetilde{G}_\tau(\boldsymbol{\eta})$ denote the same quantity but with $\mathbf{y}_t$ replaced by $\mathbf{y}_t^c = (\mathbf{x}_t, \mathbf{z}_{t-1}^c)$ or $\widetilde{\mathbf{y}}_t(\tau_t) = (\widetilde{\mathbf{x}}_t(\tau_t), \widetilde{\mathbf{z}}_{t-1}(\tau_t))$.

In addition, let

$$\begin{aligned}
\widetilde{B}_\tau(\boldsymbol{\eta}) &:= \frac{1}{Th} \sum_{t=1}^{T} \widehat{K}\left(\frac{\tau_t - \tau}{h}\right) g(\widetilde{\mathbf{y}}_t(\tau_t), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h), \\
\mathbf{m}_t^{(2)}(u, \tau) &:= \widehat{K}\left(\frac{\tau - u}{h}\right) g(\widetilde{\mathbf{y}}_t(u), \boldsymbol{\theta}(\tau) - v\mathbf{d}(u, \tau)) \cdot \mathbf{d}(u, \tau),
\end{aligned} \tag{B.2}$$

where $\mathbf{d}(u, \tau) := \boldsymbol{\theta}(u) - \boldsymbol{\theta}(\tau) - (u - \tau)\boldsymbol{\theta}^{(1)}(\tau)$ and some $v \in [0, 1]$.

Lemma B.1. *Suppose Assumptions 1 and 3 hold. Then,* $E\left(\ell(\widetilde{\mathbf{x}}_1(\tau), \widetilde{\mathbf{z}}_0(\tau); \boldsymbol{\vartheta})\right)$ *is uniquely maximized at* $\boldsymbol{\theta}(\tau)$.

Lemma B.2. *Suppose Assumptions 3–4 hold. Then,* $\ell, \boldsymbol{\nabla}\ell, \boldsymbol{\nabla}^2\ell \in \mathscr{H}(3, \boldsymbol{\chi}, M)$ *for some* $M > 0$ *and* $\boldsymbol{\chi} = \{\chi_j\}_{j=1,2,\ldots}$ *with* $\chi_j = O(j^{-(2+s)})$ *and* $s > 0$. *In addition, if* $\mathbf{H}(\mathbf{z}; \boldsymbol{\vartheta}) \equiv \mathbf{H}(\mathbf{0}; \boldsymbol{\vartheta})$, $\ell, \boldsymbol{\nabla}\ell, \boldsymbol{\nabla}^2\ell \in \mathscr{H}(2, \boldsymbol{\chi}, M)$.

Lemma B.3. *Suppose Assumptions 1–2 hold with* $r \geq C$. *Then*

1. $\sup_{\tau \in [0,1]} \left\| \sup_{\boldsymbol{\eta} \neq \boldsymbol{\eta}'} \frac{|\widetilde{G}_\tau(\boldsymbol{\eta}) - \widetilde{G}_\tau(\boldsymbol{\eta}')|}{|\boldsymbol{\eta} - \boldsymbol{\eta}'|} \right\|_1 \leq M$ *and* $\left\| \sup_{\tau \neq \tau', \boldsymbol{\eta} \neq \boldsymbol{\eta}'} \frac{|\widetilde{G}_\tau(\boldsymbol{\eta}) - \widetilde{G}_{\tau'}(\boldsymbol{\eta}')|}{|\tau - \tau'| + |\boldsymbol{\eta} - \boldsymbol{\eta}'|} \right\|_1 \leq M h^{-2}$;
2. $\sup_{\tau \in [0,1], \boldsymbol{\eta} \in \mathbf{E}_T(r)} |E(\widetilde{B}_\tau(\boldsymbol{\eta})) - \int_{-\tau/h}^{(1-\tau)/h} \widehat{K}(u) E(g(\widetilde{\mathbf{y}}_0(\tau), \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 u)) \mathrm{d}u| = O((Th)^{-1} + h)$;

3. $\left\|\sup_{\tau\neq\tau'}\frac{|\mathbf{\Pi}(\tau)-\mathbf{\Pi}(\tau')|}{|\tau-\tau'|}\right\|_1 \le Mh^{-2}$ *with* $\mathbf{\Pi}(\tau) := (Th)^{-1}\sum_{t=1}^{T}\left[\mathbf{m}_t^{(2)}(\tau,\tau_t) - E(\mathbf{m}_t^{(2)}(\tau,\tau_t))\right]$.

In addition, suppose $\chi_j = O(j^{-(2+s)})$ *for some* $s>0$*, then*

4. $\|\sup_{\tau\in[0,1],\boldsymbol{\eta}\in\mathbf{E}_T(r)}|\widetilde{G}_\tau(\boldsymbol{\eta}) - G_\tau^c(\boldsymbol{\eta})|\|_1 = O((Th)^{-1})$.

Lemma B.4. *Let* $g(\cdot)\in\mathscr{H}(C,\boldsymbol{\chi},M)$*, where* $\chi_j = O(j^{-(a+s)})$ *for some* $s>0$ *and* $a\ge 1$*. Suppose Assumptions 1–2 hold with* $q=r/C\ge 1$*, and*

$$\sup_{\tau\in[0,1]}[\alpha_j(\boldsymbol{\theta}(\tau))+\beta_j(\boldsymbol{\theta}(\tau))] = O(j^{-(a+1+s)}).$$

Then we obtain

1. $\sup_{\tau\in[0,1]}\delta_q^{\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta})|}(j) = O(j^{-(a+s)})$;
2. $\sup_{\tau\in[0,1]}\sup_{u,\boldsymbol{\eta}}\delta_q^{m(\tau,\boldsymbol{\eta},u)}(j) = O(j^{-(a+s)})$ *and* $\sup_{\tau\in[0,1]}\delta_q^{\sup_{u,\boldsymbol{\eta}}|m(\tau,\boldsymbol{\eta},u)|}(j) = O(j^{-(a+s)})$*, where* $m_t(\tau,\boldsymbol{\eta},u) := \widehat{K}\left(\frac{\tau-u}{h}\right)g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau-u)/h)$;
3. $\sup_{\tau,u\in[0,1]}\delta_q^{\mathbf{m}_t^{(2)}(u,\tau)}(j) = O(h^2j^{-(a+s)})$ *and* $\sup_{u\in[0,1]}\delta_q^{\sup_\tau|\mathbf{m}_t^{(2)}(u,\tau)|}(j) = O(h^2j^{-(a+s)})$.

Lemma B.5. *Under the conditions of Lemma B.4 with* $q=r/C>1$*, then*

1. $\|\widetilde{G}_\tau(\boldsymbol{\eta})\|_q = O\left((Th)^{-(q'-1)/q'}\right)$ *with* $q'=\min(2,q)$,
2. $\sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)}|\widetilde{G}_\tau(\boldsymbol{\eta})| = o_P(1)$;

Suppose further $q=r/C>2$ *and* $a\ge 3/2$*. Then*

3. $\sup_{\tau\in[0,1]}\sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)}|\widetilde{G}_\tau(\boldsymbol{\eta})| = O_P((\log T)^{1/2}(Th)^{-1/2}h^{-1/2})$.

Lemma B.6. *Suppose Assumptions 1–5 hold with* $r>6$*, and*

$$\sup_{\tau\in[0,1]}[\alpha_j(\boldsymbol{\theta}(\tau))+\beta_j(\boldsymbol{\theta}(\tau))] = O(j^{-(a+1+s)})$$

for $a\ge 3/2$ *and some* $s>0$*. Then*

$$\sup_{\tau\in[0,1]}\Big|\boldsymbol{\nabla}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\theta}(\tau),h\boldsymbol{\theta}^{(1)}(\tau)) - E[\boldsymbol{\nabla}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\theta}(\tau),h\boldsymbol{\theta}^{(1)}(\tau))]$$
$$-\frac{1}{Th}\sum_{t=1}^{T}\widehat{\boldsymbol{K}}((\tau_t-\tau)/h)\otimes\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))\Big| = O_P(h^2\beta_T),$$

where

$$\begin{aligned}\beta_T &= (\log T)^{1/2}(Th)^{-1/2}h^{-1/2},\\ \widehat{\boldsymbol{K}}((\tau_t-\tau)/h) &= K((\tau_t-\tau)/h)[1,(\tau_t-\tau)/h]^\top,\\ \widetilde{\mathscr{L}}_\tau(\boldsymbol{\theta}(\tau),h\boldsymbol{\theta}^{(1)}(\tau)) &:= T^{-1}\sum_{t=1}^{T}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t),\widetilde{\boldsymbol{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau)+\boldsymbol{\theta}^{(1)}(\tau)(\tau_t-\tau))K_h(\tau_t-\tau).\end{aligned}$$

Lemma B.7. *Under the conditions of Theorem 2.2,*

$$(1).\quad \sup_{\tau\in[h,1-h]}\left|-\boldsymbol{\Sigma}(\tau)(\widehat{\boldsymbol{\theta}}(\tau)-\boldsymbol{\theta}(\tau)) - \boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\mathscr{L}_\tau(\boldsymbol{\theta}(\tau),h\boldsymbol{\theta}^{(1)}(\tau))\right| = O_P(\gamma_T),$$

$$(2).\quad \sup_{\tau\in[h,1-h]}\Big|\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\mathscr{L}_\tau(\boldsymbol{\theta}(\tau),h\boldsymbol{\theta}^{(1)}(\tau)) + \frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\Sigma}(\tau)\boldsymbol{\theta}^{(2)}(\tau)$$
$$-\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t),\widetilde{\mathbf{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))K_h(\tau_t-\tau)\Big| = O_P(\beta_Th^2+h^3+(Th)^{-1}),$$

where $\beta_T = (Th)^{-1/2}h^{-1/2}(\log T)^{1/2}$ *and* $\gamma_T = (\beta_T+h)((Th)^{-1/2}\log T + h^2)$.

B.2 Secondary Lemmas

Before proceeding further, we introduce some extra notations. Assume that there exists some measurable function $\widetilde{\mathbf{H}}(\cdot,\cdot)$ such that for $\forall\tau \in [0,1]$, $\widetilde{\mathbf{h}}_t(\tau) = \widetilde{\mathbf{H}}(\tau, \mathscr{F}_t) \in \mathbb{R}^d$ is well defined, where $\mathscr{F}_t = \sigma(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \ldots)$. Let

$$\widetilde{\mathbf{D}}_{\widetilde{\mathbf{h}}}(\tau) := (Th)^{-1} \sum_{t=1}^{T} \widetilde{\mathbf{h}}_t(\tau_t) \widehat{K}((\tau_t - \tau)/h) \quad \text{and} \quad \boldsymbol{\Sigma}_{\widetilde{\mathbf{h}}}(\tau) = \sum_{j=-\infty}^{\infty} E[\widetilde{\mathbf{h}}_0(\tau)\widetilde{\mathbf{h}}_j^\top(\tau)].$$

Assume that $\boldsymbol{\Sigma}_{\widetilde{\mathbf{h}}}(\tau)$ is Lipschitz continuous and its smallest eigenvalue is bounded away from 0 uniformly over $\tau \in [0,1]$. In what follows, we let $\widetilde{h}_{0,i}(\tau)$ stand for the i^{th} component of $\widetilde{\mathbf{h}}_t(\tau)$.

Lemma B.8. *Let $q > 0$. Let $g \in \mathscr{H}(C, \boldsymbol{\chi}, M)$. Let $\mathbf{y} = (\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2, \ldots)$ and $\mathbf{y}' = (\mathbf{y}_0', \mathbf{y}_1', \mathbf{y}_2', \ldots)$ be two sequences of random variables. Assume that $\max_{j\geq 0} \|\mathbf{y}_j\|_{qC} \leq M$ and $\max_{j\geq 0} \|\mathbf{y}_j'\|_{qC} \leq M$. Then, we have*

1. $\left\|\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_r} |g(\mathbf{y},\boldsymbol{\vartheta}) - g(\mathbf{y}',\boldsymbol{\vartheta})|\right\|_q \leq M \sum_{j=0}^{\infty} \chi_j \|\mathbf{y}_j - \mathbf{y}_j'\|_{qC}$;
2. $\left\|\sup_{\boldsymbol{\vartheta}\neq\boldsymbol{\vartheta}'} \frac{|g(\mathbf{y},\boldsymbol{\vartheta}) - g(\mathbf{y},\boldsymbol{\vartheta}')|}{|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'|}\right\|_q \leq M$;
3. $\left\|\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_r} |g(\mathbf{y},\boldsymbol{\vartheta})|\right\|_q \leq M$.

Lemma B.9. *Assume that for $i = 1, \ldots, d$*

1. $\sup_{\tau\in[0,1]} \|\widetilde{h}_{0,i}(\tau)\|_q < \infty$ *with some* $2 \leq q \leq 4$,
2. $\sup_{\tau\neq\tau'} \|\widetilde{h}_{0,i}(\tau) - \widetilde{h}_{0,i}(\tau')\|_2/|\tau-\tau'| < \infty$,
3. $\sup_{\tau\in[0,1]} \delta_q^{\widetilde{h}_{0,i}(\tau)}(j) = O(j^{-(2+s)})$ *for some* $s \geq 0$.

Let $\mathbf{S}_{\widetilde{\mathbf{h}}}(t) = \sum_{s=1}^{t} \widetilde{\mathbf{h}}_s(\tau_s)$. Then on a richer probability space, there exists i.i.d. k-dimensional standard normal variables $\mathbf{v}_1, \mathbf{v}_2, \ldots$ and a process $\mathbf{S}_{\widetilde{\mathbf{h}}}^0(t) = \sum_{s=1}^{t} \boldsymbol{\Sigma}_{\widetilde{\mathbf{h}}}^{1/2}(\tau_s)\mathbf{v}_s$ such that

$$(\mathbf{S}_{\widetilde{\mathbf{h}}}(t))_{t=1}^T =_D (\mathbf{S}_{\widetilde{\mathbf{h}}}^0(t))_{t=1}^T \quad \text{and} \quad \max_{t\geq 1} |\mathbf{S}_{\widetilde{\mathbf{h}}}(t) - \mathbf{S}_{\widetilde{\mathbf{h}}}^0(t)| = O_P(\pi_T),$$

where $\pi_T = T^{\frac{q(s+3)-4}{2q(2s+3)-4}} (\log T)^{\frac{2(s+1)(q+1)}{q(2s+3)-2}}$.

Lemma B.9 is from Theorem 1 and Corollary 2 of Wu and Zhou (2011).

Lemma B.10. *Assume that for $i = 1, \ldots, d$*

1. $\sup_{\tau\in[0,1]} \|\widetilde{h}_{0,i}(\tau)\|_q < \infty$ *with some* $2 \leq q \leq 4$,
2. $\sup_{\tau\neq\tau'} \|\widetilde{h}_{0,i}(\tau) - \widetilde{h}_{0,i}(\tau')\|_2/|\tau-\tau'| < \infty$,
3. $\sup_{\tau\in[0,1]} \delta_q^{\widetilde{h}_{0,i}(\tau)}(j) = O(j^{-(2+s)})$ *for some* $s \geq 0$.

In addition, assume that $h \log T \to 0$ and $\frac{(\log T)^4}{T^{(sq+2)/(2sq+3q-2)}h} \to 0$. Then

$$\lim_{T\to\infty} \Pr\left(\sqrt{\frac{Th}{\widetilde{v}_0}} \sup_{\tau\in[h,1-h]} \left|\boldsymbol{\Sigma}_{\widetilde{\mathbf{h}}}^{-1/2}(\tau)\widetilde{\mathbf{D}}_{\widetilde{\mathbf{h}}}(\tau)\right| - B(m^*) \leq \frac{u}{\sqrt{2\log(m^*)}}\right) = \exp(-2\exp(-u)),$$

where

$$B(m^*) = \sqrt{2\log(m^*)} + \frac{\log(C_K) + (k/2 - 1/2)\log(\log(m^*)) - \log(2)}{\sqrt{2\log(m^*)}},$$

$$C_K = \frac{\{\int_{-1}^{1} |K^{(1)}(u)|^2 \mathrm{d}u/\widetilde{v}_0\pi\}^{1/2}}{\Gamma(k/2)}, \quad m^* = 1/h,$$

and $\Gamma(\cdot)$ is the Gamma function.

B.3 Proofs of Preliminary Lemmas

Proof of Lemma B.1.

Let

$$\mathbf{M}_t(\boldsymbol{\vartheta},\boldsymbol{\theta}(\tau)) := (\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top)^{-1/2}\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))^\top \times(\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top)^{-1/2}.$$

By Assumption 1.1 and the construction of $\widetilde{\mathbf{x}}_t(\tau)$, we write

$$\begin{aligned}
& E\left(\ell(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\right)\\
=& -\frac{1}{2}E\log\det\left\{\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top\right\}\\
& -\frac{1}{2}E\mathrm{tr}\left\{(\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top)^{-1}[\widetilde{\mathbf{x}}_t(\tau)-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})][\widetilde{\mathbf{x}}_t(\tau)-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})]^\top\right\}\\
=& -\frac{1}{2}E\log\det\left\{\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top\right\}-\frac{1}{2}E\mathrm{tr}\left\{\mathbf{M}_t(\boldsymbol{\vartheta},\boldsymbol{\theta}(\tau))\right\}\\
& -\frac{1}{2}E\left([\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})]^\top(\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top)^{-1}\right.\\
& \left.\times[\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})]\right)\\
=& -\frac{1}{2}\left[-E\log\det\left(\mathbf{M}_t(\boldsymbol{\vartheta},\boldsymbol{\theta}(\tau))\right)+E\mathrm{tr}\left\{\mathbf{M}_t(\boldsymbol{\vartheta},\boldsymbol{\theta}(\tau))\right\}\right]\\
& -\frac{1}{2}E\log\det\left(\widetilde{\mathbf{H}}_t(\tau,\boldsymbol{\theta}(\tau))\widetilde{\mathbf{H}}_t(\tau,\boldsymbol{\theta}(\tau))^\top\right)\\
& -\frac{1}{2}E\left([\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})]^\top(\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top)^{-1}\right.\\
& \left.\times[\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})]\right).
\end{aligned}$$

For any positive definite matrix $\mathbf{M}$ with eigenvalues $\lambda_1,\ldots,\lambda_m>0$, we have

$$f(\mathbf{M}) := -\log\det(\mathbf{M})+\mathrm{tr}\{\mathbf{M}\}=\sum_{i=1}^m(\lambda_i-\log\lambda_i)\geq m,$$

where the equality holds if $\lambda_1=\cdots=\lambda_m=1$ in which case $\mathbf{M}=\mathbf{I}_m$. Thus, $f(\mathbf{M})$ is uniquely minimized at $\mathbf{M}=\mathbf{I}_m$, which implies that $E[f(\mathbf{M}_t(\boldsymbol{\vartheta},\boldsymbol{\theta}(\tau)))]$ is uniquely minimized at $\boldsymbol{\vartheta}=\boldsymbol{\theta}(\tau)$ by Assumption 3.2. In addition, since $\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top$ is a positive definite matrix, then

$$\begin{aligned}
& E\left([\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})]^\top(\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\mathbf{H}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})^\top)^{-1}\right.\\
& \left.\times[\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\theta}(\tau))-\boldsymbol{\mu}(\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})]\right)\geq 0
\end{aligned}$$

is uniquely minimized at $\boldsymbol{\vartheta}=\boldsymbol{\theta}(\tau)$ by Assumption 3.2. Hence, $E\left(\ell(\widetilde{\mathbf{x}}_t(\tau),\widetilde{\mathbf{z}}_{t-1}(\tau);\boldsymbol{\vartheta})\right)$ is uniquely maximized at $\boldsymbol{\theta}(\tau)$. □

Proof of Lemma B.2.

We first consider $\ell(\cdot)$. Write

$$\begin{aligned}
& \ell(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta})-\ell(\mathbf{x}',\mathbf{z}';\boldsymbol{\vartheta})\\
=& -\frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))^\top\mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta})(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))-(\mathbf{x}'-\boldsymbol{\mu}(\mathbf{z}';\boldsymbol{\vartheta}))^\top\mathbf{M}^{-1}(\mathbf{z}';\boldsymbol{\vartheta})(\mathbf{x}'-\boldsymbol{\mu}(\mathbf{z}';\boldsymbol{\vartheta}))\right]\\
& -\frac{1}{2}\left[\log\det(\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta}))-\log\det(\mathbf{M}(\mathbf{z}';\boldsymbol{\vartheta}))\right]\\
:=& -\frac{1}{2}(I_1+I_2),
\end{aligned}$$

where the definitions of I_1 and I_2 should be obvious, and $\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})=\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})^\top$.

For I_2, we have

$$\begin{aligned}
|\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})-\mathbf{M}(\mathbf{z}';\boldsymbol{\vartheta})| &\leq |\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})-\mathbf{H}(\mathbf{z}';\boldsymbol{\vartheta})|\left(|\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})|+|\mathbf{H}(\mathbf{z}';\boldsymbol{\vartheta})|\right)\\
&\leq M|\mathbf{z}-\mathbf{z}'|_{\boldsymbol{\chi}}(2+|\mathbf{z}|_{\boldsymbol{\chi}}+|\mathbf{z}'|_{\boldsymbol{\chi}}),
\end{aligned}$$

where the second inequality follows from the facts that

$$|\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})-\mathbf{H}(\mathbf{z}';\boldsymbol{\vartheta})|=O(|\mathbf{z}-\mathbf{z}'|_{\boldsymbol{\chi}}),$$
$$|\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})|\leq|\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})-\mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})|+|\mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})|=O(1+|\mathbf{z}|_{\boldsymbol{\chi}})$$

by using Assumption 1.2 twice.

By Assumption 3, it is easy to know that $\det(\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta}))\geq \underline{H}>0$, which in connection with the fact $\log(\cdot)$ is Lipschitz continuous on $[\underline{H},\infty)$ yields that

$$I_2\leq M|\det(\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta}))-\det(\mathbf{M}(\mathbf{z}';\boldsymbol{\vartheta}))|.$$

In addition, for an invertible matrix $\mathbf{A}$, $\det(\mathbf{A}+\mathbf{B})=\det(\mathbf{A})+\mathrm{tr}(\mathbf{A}^{-1,\top}\mathbf{B})+o(|\mathbf{B}|)$, and for a positive definite matrix $\mathbf{A}$ and symmetric matrix $\mathbf{B}$, $|\mathrm{tr}(\mathbf{A}^{-1,\top}\mathbf{B})|\leq|\mathbf{B}|\mathrm{tr}(\mathbf{A}^{-1})$. Hence, we have

$$\begin{aligned}I_2&\leq M|\det(\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta}))-\det(\mathbf{M}(\mathbf{z}';\boldsymbol{\vartheta}))|\\&\leq M\mathrm{tr}(\mathbf{M}^{-1}(\mathbf{z}';\boldsymbol{\vartheta}))\cdot|\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta})-\mathbf{M}(\mathbf{z}';\boldsymbol{\vartheta})|\\&=O(|\mathbf{z}-\mathbf{z}'|_{\boldsymbol{\chi}}(1+|\mathbf{z}|_{\boldsymbol{\chi}}+|\mathbf{z}'|_{\boldsymbol{\chi}})).\end{aligned}$$

Note that if $\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})\equiv\mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})$, $I_2=0$.

For I_1, since $|\mathbf{H}^{-1}(\mathbf{z};\boldsymbol{\vartheta})|$ is bounded by Assumption 3, we can obtain that

$$\begin{aligned}I_1&\leq|\mathbf{H}^{-1}(\mathbf{z};\boldsymbol{\vartheta})(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))-\mathbf{H}^{-1}(\mathbf{z}';\boldsymbol{\vartheta})(\mathbf{x}'-\boldsymbol{\mu}(\mathbf{z}';\boldsymbol{\vartheta}))|\\&\quad\cdot(|\mathbf{H}^{-1}(\mathbf{z};\boldsymbol{\vartheta})(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))|+|\mathbf{H}^{-1}(\mathbf{z}';\boldsymbol{\vartheta})(\mathbf{x}'-\boldsymbol{\mu}(\mathbf{z}';\boldsymbol{\vartheta}))|)\\&=O(|\mathbf{y}-\mathbf{y}'|_{\boldsymbol{\chi}}\cdot(1+|\mathbf{y}|_{\boldsymbol{\chi}}^2+|\mathbf{y}'|_{\boldsymbol{\chi}}^2)),\end{aligned}$$

where $\mathbf{y}=(\mathbf{x},\mathbf{z})$. Similarly, if $\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})\equiv\mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})$, $I_1=O(|\mathbf{y}-\mathbf{y}'|_{\boldsymbol{\chi}}\cdot(1+|\mathbf{y}|_{\boldsymbol{\chi}}+|\mathbf{y}'|_{\boldsymbol{\chi}}))$.

For $\ell(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta})-\ell(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta}')$, write

$$\begin{aligned}&\ell(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta})-\ell(\mathbf{x},\mathbf{z};\boldsymbol{\vartheta}')\\=&-\frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))^\top\mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta})(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}))-(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}'))^\top\mathbf{M}^{-1}(\mathbf{z};\boldsymbol{\vartheta}')(\mathbf{x}-\boldsymbol{\mu}(\mathbf{z};\boldsymbol{\vartheta}'))\right]\\&-\frac{1}{2}\left[\log\det(\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta}))-\log\det(\mathbf{M}(\mathbf{z};\boldsymbol{\vartheta}'))\right]\\:=&-\frac{1}{2}(I_3+I_4).\end{aligned}$$

Similar to the development for I_1 and I_2, we can obtain that

$$I_3=O(|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'|(1+|\mathbf{y}|_{\boldsymbol{\chi}}^3)\quad\text{and}\quad I_4=O(|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'|(1+|\mathbf{z}|_{\boldsymbol{\chi}}^2)),$$

where we again let $\mathbf{y}=(\mathbf{x},\mathbf{z})$. Also if $\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})\equiv\mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})$,

$$I_3=O(|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'|(1+|\mathbf{y}|_{\boldsymbol{\chi}}^2)\quad\text{and}\quad I_4=O(|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'|).$$

Combing the above analysis, we have shown $\ell\in\mathscr{H}(3,\boldsymbol{\chi},M)$. In addition, if $\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})\equiv\mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})$, $\ell\in\mathscr{H}(2,\boldsymbol{\chi},M)$.

Similar to the development for ℓ, we can show $\boldsymbol{\nabla}\ell$, $\boldsymbol{\nabla}^2\ell\in\mathscr{H}(3,\boldsymbol{\chi},M)$ and $\boldsymbol{\nabla}\ell,\boldsymbol{\nabla}^2\ell\in\mathscr{H}(2,\boldsymbol{\chi},M)$ if $\mathbf{H}(\mathbf{z};\boldsymbol{\vartheta})\equiv\mathbf{H}(\mathbf{0};\boldsymbol{\vartheta})$.

The proof is now complete. □

Proof of Lemma B.3.

(1). By Proposition 2.1.1, we have $\sup_{\tau\in[0,1]}\|\widetilde{\mathbf{x}}_t(\tau)\|_C<\infty$. Since $g\in\mathscr{H}(C,\boldsymbol{\chi},M)$, we have

$$\sup_{\boldsymbol{\eta}\neq\boldsymbol{\eta}'}\frac{|\widetilde{G}_\tau(\boldsymbol{\eta})-\widetilde{G}_\tau(\boldsymbol{\eta}')|}{|\boldsymbol{\eta}-\boldsymbol{\eta}'|}\leq M(Th)^{-1}\sum_{t=1}^T\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)\left[2+|\widetilde{\mathbf{y}}_t(\tau_t)|_{\boldsymbol{\chi}}^C+\||\widetilde{\mathbf{y}}_t(\tau_t)|_{\boldsymbol{\chi}}^C\|_1\right].$$

Using $(Th)^{-1}\sum_{t=1}^{T}\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)<\infty$, we have

$$\left\|\sup_{\boldsymbol{\eta}\neq\boldsymbol{\eta}'}\frac{|\widetilde{G}_\tau(\boldsymbol{\eta})-\widetilde{G}_\tau(\boldsymbol{\eta}')|}{|\boldsymbol{\eta}-\boldsymbol{\eta}'|}\right\|_1\leq M\max_t\||\widetilde{\mathbf{y}}_t(\tau_t)|_{\boldsymbol{\chi}}^C\|_1<\infty.$$

In addition, by using the Lipschitz property of $\widehat{K}(\cdot)$, we have

$$\begin{aligned}
&|\widetilde{G}_\tau(\boldsymbol{\eta})-\widetilde{G}_{\tau'}(\boldsymbol{\eta}')|\\
\leq\ &(Th)^{-1}\sum_{t=1}^{T}\left|\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)-\widehat{K}\left(\frac{\tau_t-\tau'}{h}\right)\right|\cdot\sup_{\boldsymbol{\vartheta}}(|g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\vartheta})|+\|g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\vartheta})\|_1)\\
&+(Th)^{-1}\sum_{t=1}^{T}\widehat{K}\left(\frac{\tau_t-\tau'}{h}\right)\cdot|g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2(\tau_t-\tau)/h)-g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1'+\boldsymbol{\eta}_2'(\tau_t-\tau')/h)|\\
\leq\ &M\left(h^{-2}|\tau-\tau'|+h^{-1}|\boldsymbol{\eta}-\boldsymbol{\eta}'|+h^{-2}|\boldsymbol{\eta}_2|\cdot|\tau-\tau'|\right)\cdot\frac{1}{T}\sum_{t=1}^{T}(2+|\widetilde{\mathbf{y}}_t(\tau_t)|_{\boldsymbol{\chi}}^C+\||\widetilde{\mathbf{y}}_t(\tau_t)|_{\boldsymbol{\chi}}^C\|_1).
\end{aligned}$$

Combing the above analyses, the first result follows.

(2). By Lemma B.8.1 and Proposition 2.2, for $|\tau_t-\tau|\leq h$, we have

$$\begin{aligned}
&\|g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h)-g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h)\|_1\\
\leq\ &M\sum_{j=0}^{\infty}\chi_j\|\widetilde{\mathbf{x}}_{t-j}(\tau_t)-\widetilde{\mathbf{x}}_{t-j}(\tau)\|_C=O(h).
\end{aligned}$$

Hence, we have

$$\begin{aligned}
&\left\|\widetilde{B}_\tau(\boldsymbol{\eta})-\frac{1}{Th}\sum_{t=1}^{T}\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h)\right\|_1\\
\leq\ &M\frac{1}{Th}\sum_{t=1}^{T}\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)\sum_{j=0}^{\infty}\chi_j\|\widetilde{\mathbf{x}}_{t-j}(\tau_t)-\widetilde{\mathbf{x}}_{t-j}(\tau)\|_C=O(h)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{Th}\sum_{t=1}^{T}\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)E\left[g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h)\right]\\
=\ &\int_{-\tau/h}^{(1-\tau)/h}\widehat{K}(u)E(g(\widetilde{\mathbf{y}}_0(\tau),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2u))\mathrm{d}u+O((Th)^{-1})
\end{aligned}$$

by the definition of Riemann integral and the stationarity of $\widetilde{\mathbf{y}}_t(\tau)$.

(3). Write

$$\begin{aligned}
&|\mathbf{m}_t^{(2)}(u,\tau)-\mathbf{m}_t^{(2)}(u,\tau')|\\
\leq\ &|\widehat{K}\left(\frac{\tau-u}{h}\right)-\widehat{K}\left(\frac{\tau'-u}{h}\right)|\cdot|g(\widetilde{\mathbf{y}}_t(u),\boldsymbol{\theta}(\tau)-v\mathbf{d}(u,\tau))|\cdot|\mathbf{d}(u,\tau)|\\
&+|\widehat{K}\left(\frac{\tau'-u}{h}\right)|\cdot|g(\widetilde{\mathbf{y}}_t(u),\boldsymbol{\theta}(\tau)-v\mathbf{d}(u,\tau))-g(\widetilde{\mathbf{y}}_t(u),\boldsymbol{\theta}(\tau')-v\mathbf{d}(u,\tau'))|\cdot|\mathbf{d}(u,\tau)|\\
&+|\widehat{K}\left(\frac{\tau'-u}{h}\right)|\cdot|g(\widetilde{\mathbf{y}}_t(u),\boldsymbol{\theta}(\tau')-v\mathbf{d}(u,\tau'))|\cdot|\mathbf{d}(u,\tau)-\mathbf{d}(u,\tau')|\\
:=\ &I_1+I_2+I_3.
\end{aligned}$$

By the Lipschitz continuity of $\widehat{K}(\cdot)$ and $\|\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(u),\boldsymbol{\vartheta})|\|_1=O(1)$ (by Lemma B.8.3), we have

$$E(I_1)=O(h^{-1}|\tau-\tau'|).$$

Similarly, by Lemma B.8.2 and $|\mathbf{d}(u,\tau)| = O(1)$, we have

$$E(I_2) = O(|\tau-\tau'|).$$

By the Lipschitz continuity of $\mathbf{d}(u,\cdot)$, we have $E(I_3) = O(|\tau-\tau'|)$. Hence,

$$\left|\frac{1}{Th}\sum_{t=1}^{T}[\mathbf{m}_t^{(2)}(\tau,\tau_t) - \mathbf{m}_t^{(2)}(\tau',\tau_t)]\right| \leq \frac{1}{Th}\sum_{t=1}^{T}|\mathbf{m}_t^{(2)}(\tau,\tau_t) - \mathbf{m}_t^{(2)}(\tau',\tau_t)| = O(h^{-2}|\tau-\tau'|).$$

The proof is now complete.

(4). By Propositions 2.1.1 and 2.2.2, we have $\sup_{\tau\in[0,1]}\|\widetilde{\mathbf{x}}_t(\tau)\|_C < \infty$ and $\max_t\|\mathbf{x}_t - \widetilde{\mathbf{x}}_t(\tau_t)\|_C = O(T^{-1})$. Hence, we have $\max_t\|\mathbf{x}_t\|_C \leq M$.

By Lemma B.8 and the definitions of $\mathbf{y}_t$ and $\mathbf{y}_t^c$, we have

$$\|\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_r}|g(\mathbf{y}_t,\boldsymbol{\vartheta}) - g(\mathbf{y}_t^c,\boldsymbol{\vartheta})|\|_1 \leq M\sum_{j=t}^{\infty}\chi_j\|\mathbf{x}_{t-j}\|_C = O\left(\sum_{j=t}^{\infty}\chi_j\right).$$

In addition, by Proposition 2.2.1 and $\chi_j = O(j^{-(2+s)})$ for some $s>0$, we have

$$\begin{aligned}
&\|\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_r}|g(\mathbf{y}_t,\boldsymbol{\vartheta}) - g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\vartheta})|\|_1 \leq M\sum_{j=0}^{\infty}\chi_j\|\mathbf{x}_{t-j} - \widetilde{\mathbf{x}}_{t-j}(\tau_t)\|_C \\
&\leq M\sum_{j=0}^{\infty}\chi_j\|\mathbf{x}_{t-j} - \widetilde{\mathbf{x}}_{t-j}(\tau_{t-j})\|_C + M\sum_{j=0}^{\infty}\chi_j\|\widetilde{\mathbf{x}}_{t-j}(\tau_t) - \widetilde{\mathbf{x}}_{t-j}(\tau_{t-j})\|_C \\
&= O(\sum_{j=0}^{\infty}\chi_j/T) + O(\sum_{j=0}^{\infty}j\chi_j/T) = O(T^{-1}).
\end{aligned}$$

Hence, we have

$$\begin{aligned}
&\|\sup_{\tau\in[0,1]}\sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)}|\widetilde{G}_\tau(\boldsymbol{\eta}) - G_\tau^c(\boldsymbol{\eta})|\|_1 \\
&\leq M(Th)^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}_r}\|g(\widetilde{\mathbf{y}}_t,\boldsymbol{\vartheta}) - g(\mathbf{y}_t^c,\boldsymbol{\vartheta})\|_1 \\
&\leq M(Th)^{-1}\sum_{t=1}^{T}\sum_{j=t}^{\infty}\chi_j \leq M(Th)^{-1}\sum_{j=1}^{\infty}j\chi_j = O((Th)^{-1}).
\end{aligned}$$

The proof of the fourth result is now complete. □

Proof of Lemma B.4.

(1). Let $\widetilde{\mathbf{y}}_t^*(\tau)$ be a coupled version of $\widetilde{\mathbf{y}}_t(\tau)$ with $\boldsymbol{\varepsilon}_0$ replaced by $\boldsymbol{\varepsilon}_0^*$. By Lemma B.8, we have

$$\begin{aligned}
\delta_q^{\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta})|}(t) &= \|\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta})| - \sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t^*(\tau),\boldsymbol{\vartheta})|\|_q \\
&\leq \|\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta}) - g(\widetilde{\mathbf{y}}_t^*(\tau),\boldsymbol{\vartheta})|\|_q \\
&\leq M\sum_{j=0}^{\infty}\chi_j\|\widetilde{\mathbf{x}}_{t-j}(\tau) - \widetilde{\mathbf{x}}_{t-j}^*(\tau)\|_{qC} \\
&= M\sum_{j=0}^{t}\chi_j\delta_r^{\widetilde{\mathbf{x}}(\tau)}(t-j).
\end{aligned}$$

By Proposition 2.1.2 and the conditions on $\alpha_j(\boldsymbol{\theta}(\tau))$ and $\beta_j(\boldsymbol{\theta}(\tau))$ in the body of this lemma, we have $\delta_r^{\widetilde{\mathbf{x}}(\tau)}(j) = O(j^{-(a+s)})$ for some $s>0$. Hence, we have

$$\sum_{j=0}^{t}\chi_j\delta_r^{\widetilde{\mathbf{x}}(\tau)}(t-j) \leq \sum_{j\geq t/2}\chi_j\delta_r^{\widetilde{\mathbf{x}}(\tau)}(t-j) + \sum_{0\leq j\leq t/2}\chi_j\delta_r^{\widetilde{\mathbf{x}}(\tau)}(t-j)$$

$$
\begin{aligned}
&\le (t/2)^{-(a+s)} \sum_{j\ge t/2} \delta_r^{\widetilde{\mathbf{x}}(\tau)}(t-j) + (t/2)^{-(a+s)} \sum_{0\le j\le t/2} \chi_j \\
&= O(t^{-(a+s)}).
\end{aligned}
$$

The proof of the first result of this lemma is now complete.
(2)–(3). Since

$$
\begin{aligned}
|\sup_{u,\boldsymbol{\eta}} |m(\tau,\boldsymbol{\eta},u)| - \sup_{u,\boldsymbol{\eta}} |m^*(\tau,\boldsymbol{\eta},u)|| &\le \sup_{u,\boldsymbol{\eta}} |m(\tau,\boldsymbol{\eta},u) - m^*(\tau,\boldsymbol{\eta},u)| \\
&\le M \sup_{\boldsymbol{\vartheta}} |g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta}) - g(\widetilde{\mathbf{y}}_t^*(\tau),\boldsymbol{\vartheta})|,
\end{aligned}
$$

the second result follows directly from the first result.

Since $\mathbf{d}(u,\tau) = O(h^2)$ when $|\tau - u| \le h$, for each element of $\mathbf{m}_t^{(2)}(u,\tau)$, we have

$$
\begin{aligned}
|\sup_{\tau} |m_{t,i}^{(2)}(u,\tau)| - \sup_{\tau} |m_{t,i}^{(2)}(\tau,u)^*|| &\le \sup_{\tau} |m_{t,i}^{(2)}(\tau,u) - m_{t,i}^{(2)}(\tau,u)^*| \\
&\le Mh^2 \sup_{\boldsymbol{\vartheta}} |g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta}) - g(\widetilde{\mathbf{y}}_t^*(\tau),\boldsymbol{\vartheta})|,
\end{aligned}
$$

where $m_{t,i}^{(2)}(\tau,u)$ is yielded by the coupled version.

The proof is now complete. □

Proof of Lemma B.5.

(1). Note that

$$
\begin{aligned}
&g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) - E(g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h)) \\
&= \sum_{l=0}^{\infty} \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h),
\end{aligned}
$$

in which $\{\mathscr{P}_{t-l}(g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h))\}_{t=1}^T$ is a sequence of martingale differences.

If $1 < q \le 2$, by the Burkholder inequality, $|\sum_{i=1}^d a_i|^r \le \sum_{i=1}^d |a_i|^r$ for $r \in (0,1]$ and Lemma B.4.1, we have

$$
\begin{aligned}
\|\widetilde{G}_\tau(\boldsymbol{\eta})\|_q &\le \sum_{l=0}^{\infty} \left\| \sum_{t=1}^T \frac{1}{Th}\widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) \right\|_q \\
&\le O(1) \sum_{l=0}^{\infty} \left\{ E\left[\sum_{t=1}^T \left(\frac{1}{Th}\widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) \right)^2 \right]^{q/2} \right\}^{1/q} \\
&\le O(1) \sum_{l=0}^{\infty} \left\{ E\left[\sum_{t=1}^T \left(\frac{1}{Th}\widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) \right)^q \right] \right\}^{1/q} \\
&\le O(1)(Th)^{-(q-1)/q} \sum_{l=0}^{\infty} \sup_{\tau\in[0,1]} \delta_q^{\sup_{\boldsymbol{\vartheta}} |g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\vartheta})|}(l) \left(\frac{1}{Th} \sum_{t=1}^T \widehat{K}\left(\frac{\tau_t - \tau}{h}\right)^q \right)^{1/q} \\
&= O((Th)^{-(q-1)/q}).
\end{aligned}
$$

Similarly, for $q \ge 2$, by the Burkholder inequality and the Minkowski inequality, we have

$$
\begin{aligned}
\|\widetilde{G}_\tau(\boldsymbol{\eta})\|_q &= \left\| \frac{1}{Th} \sum_{t=1}^T \widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \sum_{l=0}^{\infty} \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) \right\|_q \\
&\le \sum_{l=0}^{\infty} \left\| \sum_{t=1}^T \frac{1}{Th}\widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) \right\|_q \\
&\le O(1) \sum_{l=0}^{\infty} \left\{ E\left[\sum_{t=1}^T \left(\frac{1}{Th}\widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) \right)^2 \right]^{q/2} \right\}^{1/q} \\
&\le O(1) \sum_{l=0}^{\infty} \left\{ \sum_{t=1}^T \left[E\left(\frac{1}{Th}\widehat{K}\left(\frac{\tau_t - \tau}{h}\right) \mathscr{P}_{t-l} g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) \right)^q \right]^{2/q} \right\}^{1/2}
\end{aligned}
$$

$$= O(1)(Th)^{-1/2}\sum_{l=0}^{\infty}\sup_{\tau\in[0,1]}\delta_q^{\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta})|}(l) = O((Th)^{-1/2}).$$

The proof of the first result is now complete.

(2). For any fixed $v>0$, let $\kappa>0$ and $\mathbf{E}_T^{\kappa}(r)$ be a discretization of $\mathbf{E}_T(r)$ such that for each $\boldsymbol{\eta}\in\mathbf{E}_T(r)$ one can find $\boldsymbol{\eta}'\in\mathbf{E}_T^{\kappa}(r)$ satisfying $|\boldsymbol{\eta}-\boldsymbol{\eta}'|\le\kappa$. Let $\#\mathbf{E}_T^{\kappa}(r)$ denote the numbers of sets in $\mathbf{E}_T^{\kappa}(r)$. Write

$$\Pr\left(\sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)}|\widetilde{G}_\tau(\boldsymbol{\eta})|>v\right)\le\#\mathbf{E}_T^{\kappa}(r)\sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)}\Pr\Big(|\widetilde{G}_\tau(\boldsymbol{\eta})|>v/2\Big)$$
$$+\Pr\left(\sup_{|\boldsymbol{\eta}-\boldsymbol{\eta}'|\le\kappa}|\widetilde{G}_\tau(\boldsymbol{\eta})-\widetilde{G}_\tau(\boldsymbol{\eta}')|>v/2\right).$$

By the Markov inequality, we have

$$\Pr\Big(|\widetilde{G}_\tau(\boldsymbol{\eta})|>v/2\Big)\le\frac{\|\widetilde{G}_\tau(\boldsymbol{\eta})\|_q^q}{(v/2)^q}.$$

Note that $\{\mathscr{P}_{t-j}g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h)\}_t$ forms a sequence of martingale differences. By the Burkholder inequality and Lemma B.4.1, we have

$$\|\widetilde{G}_\tau(\boldsymbol{\eta})\|_q\le(Th)^{-1}\sum_{j=0}^{\infty}\|\sum_{t=1}^{T}\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)\mathscr{P}_{t-j}g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h)\|_q$$
$$\le(q-1)^{-1}(Th)^{-1}\sum_{j=0}^{\infty}\left(\|\sum_{t=1}^{T}\widehat{K}\left(\frac{\tau_t-\tau}{h}\right)^2\mathscr{P}_{t-j}^2g(\widetilde{\mathbf{y}}_t(\tau_t),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau_t-\tau)/h)\|_{q/2}^{q/2}\right)^{1/q}$$
$$\le M(Th)^{-(q-1)/q}\sum_{j=0}^{\infty}\sup_{\tau\in[0,1]}\delta_q^{\sup_{\boldsymbol{\vartheta}}|g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\vartheta})|}(j)=O((Th)^{-(q-1)/q}),$$

which in connection with the fact $\#\mathbf{E}_T^{\kappa}(r)$ is independent of T yields that

$$\#\mathbf{E}_T^{\kappa}(r)\sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)}\Pr\Big(|\widetilde{G}_\tau(\boldsymbol{\eta})|>v/2\Big)=o(1).$$

In addition, by Lemma B.3.1, we have

$$\Pr\left(\sup_{|\boldsymbol{\eta}-\boldsymbol{\eta}'|\le\kappa}|\widetilde{G}_\tau(\boldsymbol{\eta})-\widetilde{G}_\tau(\boldsymbol{\eta}')|>v/2\right)\le M\kappa\to 0$$

by choosing κ small enough. Hence, $\Pr\Big(\sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)}|\widetilde{G}_\tau(\boldsymbol{\eta})|>v\Big)\to 0$ as $T\to\infty$.

(3). Let $\beta_T:=(\log T)^{1/2}(Th)^{-1/2}h^{-1/2}$ for short. Let further $\mathbf{E}_{T,\kappa}(r)$ be a discretization of $\mathbf{E}_T(r)$ such that for each $\boldsymbol{\eta}\in\mathbf{E}_T(r)$ one can find $\boldsymbol{\eta}'\in\mathbf{E}_{T,\kappa}(r)$ satisfying $|\boldsymbol{\eta}-\boldsymbol{\eta}'|\le\kappa_T^{-1}$. Define $\mathscr{T}_{T,\kappa}=\{t/\kappa_T: t=1,2,...,\kappa_T\}$ as a discretization of $[0,1]$. For some constant $M>0$, we have

$$\Pr\left(\sup_{\tau\in[0,1]}\sup_{\boldsymbol{\eta}\in\mathbf{E}_T(r)}|\widetilde{G}_\tau(\boldsymbol{\eta})|>M\beta_T\right)$$
$$\le\Pr\left(\sup_{\tau\in\mathscr{T}_{T,\kappa}}\sup_{\boldsymbol{\eta}\in\mathbf{E}_{T,\kappa}(r)}|\widetilde{G}_\tau(\boldsymbol{\eta})|>\beta_T M/2\right)$$
$$+\Pr\left(\sup_{|\tau-\tau'|\le\kappa_T^{-1},|\boldsymbol{\eta}-\boldsymbol{\eta}'|\le\kappa_T^{-1}}|\widetilde{G}_\tau(\boldsymbol{\eta})-\widetilde{G}_{\tau'}(\boldsymbol{\eta}')|>\beta_T M/2\right).$$

Let $m_t(\tau,\boldsymbol{\eta},u):=\widehat{K}\left(\frac{\tau-u}{h}\right)g(\widetilde{\mathbf{y}}_t(\tau),\boldsymbol{\eta}_1+\boldsymbol{\eta}_2\cdot(\tau-u)/h)$, we have $\sup_{\tau\in[0,1]}\sup_{u,\boldsymbol{\eta}}\delta_q^{m(\tau,\boldsymbol{\eta},u)}(j)=O(j^{-(a+s)})$ and $\sup_{\tau\in[0,1]}\delta_q^{\sup_{u,\boldsymbol{\eta}}|m(\tau,\boldsymbol{\eta},u)|}(j)=O(j^{-(a+s)})$ for some $a\ge 3/2$ by Lemma B.4.2. Let $\alpha=1/2$, we have

$$W_{q,\alpha}:=\max_{k\ge 0}(k+1)^{\alpha}\sup_{\tau\in[0,1]}\sum_{j=k}^{\infty}\delta_q^{\sup_{u,\boldsymbol{\eta}}|m(\tau,\boldsymbol{\eta},u)|}(j)\le M\max_k k^{-(a-3/2+s)}<\infty$$

and

$$W_{2,\alpha} := \max_{k\geq 0}(k+1)^{\alpha} \sup_{\tau\in[0,1]} \sup_{u,\boldsymbol{\eta}} \sum_{j=k}^{\infty} \delta_2^{m(\tau,\boldsymbol{\eta},u)}(j) \leq M \max_{k} k^{-(a-3/2+s)} < \infty.$$

Note that $l = \min\{1, \log(\#\mathbf{E}_{T,\kappa}(r) \times \mathscr{T}_{T,\kappa})\} \leq 3(2d+1)\log(T)$ and $M\beta_T Th = MT^{1/2}(\log T)^{1/2} \geq \sqrt{Tl}W_{2,\alpha} + T^{1/q}l^{3/2}W_{q,\alpha} \geq T^{1/2}(\log T)^{1/2} + T^{1/q}(\log T)^{3/2}$ for some M large enough. By using Theorem 6.2 of Zhang and Wu (2017) (the proof therein also works for the uniform functional dependence measure) with $q > 2$ and $\alpha = 1/2$ to $\{m_t(\tau, \boldsymbol{\eta}, \tau_t)\}_{\tau\in\mathscr{T}_{T,\kappa},\boldsymbol{\eta}\in\mathbf{E}_{T,\kappa}(r)}$, we have

$$\begin{aligned}
&\Pr\left(\sup_{\tau\in\mathscr{T}_{T,\kappa}} \sup_{\boldsymbol{\eta}\in\mathbf{E}_{T,\kappa}(r)} |\widetilde{G}_\tau(\boldsymbol{\eta})| > \beta_T M/2\right)\\
&\leq \frac{MTl^{q/2}}{(\beta_T Th)^q} + M\exp\left(-\frac{M(\beta_T Th)^2}{T}\right)\\
&\leq M\left(T^{-(q-2)/2} + \exp(-\log T)\right) \to 0.
\end{aligned}$$

In addition, by the Markov inequality and Lemma B.3.1, we have

$$\Pr\left(\sup_{|\tau-\tau'|\leq\kappa_T^{-1}, |\boldsymbol{\eta}-\boldsymbol{\eta}'|\leq\kappa_T^{-1}} |\widetilde{G}_\tau(\boldsymbol{\eta}) - \widetilde{G}_{\tau'}(\boldsymbol{\eta}')| > \beta_T M/2\right) = O(h^{-2}T^{-3}/\beta_T) \to 0.$$

The proof is now complete. □

Proof of Lemma B.6.

For notational simplicity, we let $\boldsymbol{\eta}(\tau) = [\boldsymbol{\theta}(\tau), h\boldsymbol{\theta}^{(1)}(\tau)]$ in what follows, and define

$$\begin{aligned}
\boldsymbol{\Gamma}(\tau) &:= \boldsymbol{\nabla}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) - E[\boldsymbol{\nabla}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))]\\
&\quad -\frac{1}{Th}\sum_{t=1}^{T} \widehat{\boldsymbol{K}}((\tau_t-\tau)/h) \otimes \boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t)).
\end{aligned}$$

Due to $E(\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t))) = 0$ by Lemma B.1, we have

$$\begin{aligned}
\boldsymbol{\Gamma}(\tau) = &\frac{1}{Th}\sum_{t=1}^{T} \widehat{\boldsymbol{K}}((\tau_t-\tau)/h) \otimes [\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau) + \boldsymbol{\theta}^{(1)}(\tau)(\tau_t-\tau))\\
&-\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t))]\\
&-\frac{1}{Th}\sum_{t=1}^{T} \widehat{\boldsymbol{K}}((\tau_t-\tau)/h) \otimes E[\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau) + \boldsymbol{\theta}^{(1)}(\tau)(\tau_t-\tau))\\
&-\boldsymbol{\nabla}_{\boldsymbol{\vartheta}}\ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t)].
\end{aligned}$$

By the Mean Value Theorem, we have

$$\boldsymbol{\Gamma}(\tau) = (Th)^{-1}\sum_{t=1}^{T}[\mathbf{M}_t^{(2)}(\tau,\tau_t) - E(\mathbf{M}_t^{(2)}(\tau,\tau_t))],$$

where

$$\begin{aligned}
&\mathbf{M}_t^{(2)}(\tau,u) := \widehat{\boldsymbol{K}}((\tau_t-\tau)/h) \otimes \boldsymbol{\nabla}_{\boldsymbol{\vartheta}}^2\ell(\widetilde{\boldsymbol{x}}_t(u), \widetilde{\boldsymbol{z}}_{t-1}(u); \boldsymbol{\theta}(\tau) - v\mathbf{r}(u))\mathbf{r}(u) \text{ for some } v\in[0,1],\\
&\mathbf{r}(u) = \frac{1}{2}\boldsymbol{\theta}^{(2)}(\tau)(u-\tau)^2 + \frac{1}{6}\boldsymbol{\theta}^{(3)}(\overline{\tau})(u-\tau)^3 \text{ with } \overline{\tau} \text{ between } u \text{ and } \tau.
\end{aligned}$$

We then use a similar argument as in the proof of Lemma B.5 to prove

$$\Pr\left(\sup_{\tau\in[0,1]} |\boldsymbol{\Gamma}(\tau)| > M\beta_T h^2\right) \to 0.$$

Define $\kappa_T = T^5$ and $\mathscr{T}_{T,\kappa} = \{t/\kappa_T : t = 1, 2, ..., \kappa_T\}$ as a discretization of $[0,1]$. For some constant $M > 0$, we have

$$\Pr\left(\sup_{\tau\in[0,1]}|\mathbf{\Gamma}(\tau)|>M\beta_T h^2\right)\le \Pr\left(\sup_{\tau\in\mathscr{T}_{T,\kappa}}|\mathbf{\Gamma}(\tau)|>\beta_T h^2 M/2\right)$$
$$+\Pr\left(\sup_{|\tau-\tau'|\le\kappa_T^{-1}}|\mathbf{\Gamma}(\tau)-\mathbf{\Gamma}(\tau')|>\beta_T h^2 M/2\right).$$

By Lemma B.3.3 and the Markov inequality, we have

$$\Pr\left(\sup_{|\tau-\tau'|\le\kappa_T^{-1}}|\mathbf{\Gamma}(\tau)-\mathbf{\Gamma}(\tau')|>\beta_T h^2 M/2\right)=O\left(\frac{h^{-2}\kappa_T^{-1}}{\beta_T h^2 M/2}\right)\to 0.$$

By Lemma B.4.3, we have

$$\sup_{u,\tau\in[0,1]}\delta_q^{|\mathbf{M}^{(2)}(\tau,u)|}(j)=O(h^2 j^{-(a+s)})\quad\text{and}\quad\sup_{\tau\in[0,1]}\delta_q^{\sup_u|\mathbf{M}^{(2)}(\tau,u)|}(j)=O(h^2 j^{-(a+s)})$$

for some $a\ge 3/2$. Let $\alpha=1/2$, we have

$$\widetilde{W}_{q,\alpha}:=\max_{k\ge0}(k+1)^\alpha\sup_{\tau\in[0,1]}\sum_{j=k}^{\infty}\delta_q^{\sup_u|\mathbf{M}^{(2)}(\tau,u)|}(j)=O(h^2)$$

and

$$\widetilde{W}_{2,\alpha}:=\max_{k\ge0}(k+1)^\alpha\sup_{\tau,u\in[0,1]}\sum_{j=k}^{\infty}\delta_2^{|\mathbf{M}^{(2)}(\tau,u)|}(j)=O(h^2).$$

Using Theorem 6.2 of Zhang and Wu (2017) with $q>2$, $\alpha=1/2$ and $l=\min\{1,\log(\#\mathscr{T}_{T,\kappa})\}\le 5\log(T)$ to $\{\mathbf{M}_t^{(2)}(\tau,\tau_t)\}_{\tau\in\mathscr{T}_{T,\kappa}}$, we have

$$\begin{aligned}&\Pr\left(\sup_{\tau\in\mathscr{T}_{T,\kappa}}|\mathbf{\Gamma}(\tau)|>h^2\beta_T M/2\right)\\
&\le\frac{MTl^{q/2}\widetilde{W}_{q,\alpha}^q}{(\beta_T h^2 Th)^q}+M\exp\left(-\frac{M(\beta_T h^2 Th)^2}{T\widetilde{W}_{2,\alpha}^2}\right)\\
&\le M\left(T^{-(q-2)/2}+\exp(-\log T)\right)\to 0.\end{aligned}$$

The proof is now complete. ☐

Proof of Lemma B.7.

(1). Let $\widehat{\boldsymbol{\eta}}(\tau):=[\widehat{\boldsymbol{\theta}}(\tau)^\top,\widehat{\boldsymbol{\theta}}^\star(\tau)^\top]^\top$ and $\boldsymbol{\eta}(\tau):=[\boldsymbol{\theta}(\tau)^\top,h\boldsymbol{\theta}^{(1)}(\tau)^\top]^\top$. By Lemma B.5 and the proof of Theorem 2.1, we have

$$\sup_{\tau\in[0,1]}|\widehat{\boldsymbol{\eta}}(\tau)-\boldsymbol{\eta}(\tau)|=o_P(1).$$

By the Taylor expansion, we have

$$\widehat{\boldsymbol{\eta}}(\tau)-\boldsymbol{\eta}(\tau)=-(\widetilde{\mathbf{\Sigma}}(\tau)+\mathbf{R}_T(\tau))^{-1}\nabla\mathscr{L}_\tau(\boldsymbol{\eta}(\tau)),$$

where $\mathbf{R}_T(\tau):=\nabla^2\mathscr{L}_\tau(\overline{\boldsymbol{\eta}})-\widetilde{\mathbf{\Sigma}}(\tau)$ and $\widetilde{\mathbf{\Sigma}}(\tau):=\begin{bmatrix}1&0\\0&\widetilde{c}_2\end{bmatrix}\otimes\mathbf{\Sigma}(\tau)$ with $\overline{\boldsymbol{\eta}}$ between $\widehat{\boldsymbol{\eta}}(\tau)$ and $\boldsymbol{\eta}(\tau)$. By Lemma B.3 and Lemma B.5, we have

$$\begin{aligned}&\sup_{\tau\in[0,1],\boldsymbol{\eta}\in\mathbf{E}_T(r)}|\nabla^2\mathscr{L}_\tau(\boldsymbol{\eta})-\widetilde{\mathbf{\Sigma}}(\tau,\boldsymbol{\eta})|\\
=&\sup_{\tau\in[0,1],\boldsymbol{\eta}\in\mathbf{E}_T(r)}|\nabla^2\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta})-\widetilde{\mathbf{\Sigma}}(\tau,\boldsymbol{\eta})|+O_P((Th)^{-1})\\
=&\sup_{\tau\in[0,1],\boldsymbol{\eta}\in\mathbf{E}_T(r)}|E[\nabla^2\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta})]-\widetilde{\mathbf{\Sigma}}(\tau,\boldsymbol{\eta})|+O_P((Th)^{-1}+\beta_T)\end{aligned}$$

$$= O_P(\beta_T + (Th)^{-1}) + O(h),$$

where $\widetilde{\mathbf{\Sigma}}(\tau, \boldsymbol{\eta}) := \int_{-\tau/h}^{(1-\tau)/h} K(u) \begin{bmatrix} 1 & u \\ u & u^2 \end{bmatrix} \otimes \mathbf{\Sigma}(\tau, \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 u) \mathrm{d}u$ and

$$\mathbf{\Sigma}(\tau, \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 u) := E\left(\nabla^2_{\boldsymbol{\vartheta}} \ell(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 u)\right).$$

By Lemma B.4.3 and the condition

$$\sup_{\tau\in[0,1]} [\alpha_j(\boldsymbol{\theta}(\tau)) + \beta_j(\boldsymbol{\theta}(\tau))] = O(j^{-(3+s)})$$

for some $s > 0$, we have $\sup_{\tau\in[0,1]} \delta_q^{\nabla_{\vartheta_j}\ell}(j) = O(j^{-(2+s)})$ for some $s > 0$. By Lemma B.9, we have

$$\sup_{\tau\in[0,1]} \left| T^{-1} \sum_{t=1}^{T} K_h(\tau_t - \tau) \nabla_{\boldsymbol{\vartheta}} \ell(\widetilde{\mathbf{x}}_t(\tau_t), \widetilde{\mathbf{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t)) \right| = O_P((Th)^{-1/2} \log T).$$

Since $E\left(\nabla_{\boldsymbol{\vartheta}} \ell(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{z}}_{t-1}(\tau); \boldsymbol{\theta}(\tau))\right) = 0$, we further obtain that

$$\begin{aligned}
&\sup_{\tau\in[0,1]} |\nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) - E[\nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))]| \\
\leq& \sup_{\tau\in[0,1]} |\nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) - E[\nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))] \\
&- \frac{1}{Th} \sum_{t=1}^{T} \widehat{\boldsymbol{K}}((\tau_t - \tau)/h) \otimes \nabla_{\boldsymbol{\vartheta}} \ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t))| \\
&+ \sup_{\tau\in[0,1]} \left| T^{-1} \sum_{t=1}^{T} K_h(\tau_t - \tau) \nabla_{\boldsymbol{\vartheta}} \ell(\widetilde{\mathbf{x}}_t(\tau_t), \widetilde{\mathbf{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t)) \right| \\
=& O_P(h^2 \beta_T + (Th)^{-1/2} \log T)
\end{aligned}$$

using Lemma B.6.

Hence, by Lemma B.3.4, we have

$$\begin{aligned}
&\sup_{\tau\in[0,1]} |\nabla \mathscr{L}_\tau(\boldsymbol{\eta}(\tau))| \\
\leq& \sup_{\tau\in[0,1]} |\nabla \mathscr{L}_\tau(\boldsymbol{\eta}(\tau)) - \nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))| + \sup_{\tau\in[0,1]} |\nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) - E(\nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)))| \\
&+ \sup_{\tau\in[0,1]} |E(\nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)))| \\
=& \sup_{\tau\in[0,1]} |E(\nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)))| + O_P(h^2 \beta_T + (Th)^{-1/2} \log T + (Th)^{-1}).
\end{aligned}$$

Since

$$\begin{aligned}
&E\left[\nabla \widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)) - \frac{1}{Th} \sum_{t=1}^{T} \widehat{\boldsymbol{K}}((\tau_t - \tau)/h) \otimes \nabla_{\boldsymbol{\vartheta}} \ell(\widetilde{\boldsymbol{x}}_t(\tau_t), \widetilde{\boldsymbol{z}}_{t-1}(\tau_t); \boldsymbol{\theta}(\tau_t))\right] \\
=& -\frac{1}{2} h^2 \frac{1}{Th} \sum_{t=1}^{T} \widehat{\boldsymbol{K}}((\tau_t - \tau)/h) \otimes \left[E[\nabla^2_{\boldsymbol{\vartheta}} \ell(\widetilde{\boldsymbol{x}}_t(\tau), \widetilde{\boldsymbol{z}}_{t-1}(\tau); \boldsymbol{\theta}(\tau))] \boldsymbol{\theta}^{(2)}(\tau) \left(\frac{\tau_t - \tau}{h}\right)^2\right] + O(h^3) \\
=& \frac{1}{2} h^2 \int_{-\tau/h}^{(1-\tau)/h} K(u) [u^2, u^3]^\top \mathrm{d}u \otimes \left(-\mathbf{\Sigma}(\tau) \boldsymbol{\theta}^{(2)}(\tau)\right) + O((Th)^{-1} + h^3),
\end{aligned}$$

we have

$$\sup_{\tau\in[0,1]} |\nabla_{\boldsymbol{\eta}_j} \mathscr{L}_\tau(\boldsymbol{\eta}(\tau))| = O_P(h^2 \beta_T + (Th)^{-1/2} \log T + (Th)^{-1} + h^{1+j}).$$

for $j = 1, 2$. Hence, we have $\sup_{\tau\in[0,1]} |\widehat{\boldsymbol{\eta}}_j(\tau) - \boldsymbol{\eta}_j(\tau)| = O_P(h^2 \beta_T + (Th)^{-1/2} \log T + (Th)^{-1} + h^{1+j})$ and $\sup_{\tau\in[0,1]} |\mathbf{R}_T(\tau)| = O_P(\beta_T + h + (Th)^{-1})$, where $\widehat{\boldsymbol{\eta}}_j(\tau)$ and $\boldsymbol{\eta}_j(\tau)$ are corresponding to the j^{th} part in their definitions given in the beginning of this proof.

Write

$$\begin{aligned}
&\left|-\widetilde{\mathbf{\Sigma}}(\tau)(\widehat{\boldsymbol{\eta}}(\tau)-\boldsymbol{\eta}(\tau))-\nabla\mathscr{L}_\tau(\boldsymbol{\eta}(\tau))\right| \\
&\leq |[\mathbf{I}_{2d}+\widetilde{\mathbf{\Sigma}}^{-1}(\tau)\mathbf{R}_T(\tau)]^{-1}-\mathbf{I}_{2d}^{-1}|\cdot|\nabla\mathscr{L}_\tau(\boldsymbol{\eta}(\tau))| \\
&\leq |[\mathbf{I}_{2d}+\widetilde{\mathbf{\Sigma}}^{-1}(\tau)\mathbf{R}_T(\tau)]^{-1}|\cdot|\widetilde{\mathbf{\Sigma}}^{-1}(\tau)\mathbf{R}_T(\tau)|\cdot|\nabla\mathscr{L}_\tau(\boldsymbol{\eta}(\tau))| \\
&= O_P(\gamma_T).
\end{aligned}$$

The proof of the first result is now complete.
(2). By Lemma B.3 and Lemma B.6, we have

$$\begin{aligned}
&\sup_{\tau\in[h,1-h]}\Big|\nabla_{\boldsymbol{\vartheta}}\mathscr{L}_\tau(\boldsymbol{\theta}(\tau),h\boldsymbol{\theta}^{(1)}(\tau))+\frac{1}{2}h^2\widetilde{c}_2\mathbf{\Sigma}(\tau)\boldsymbol{\theta}^{(2)}(\tau) \\
&\qquad -\frac{1}{T}\sum_{t=1}^T\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t),\widetilde{\mathbf{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))K_h(\tau_t-\tau)\Big| \\
\leq &\sup_{\tau\in[h,1-h]}|\nabla_{\boldsymbol{\eta}_1}\mathscr{L}_\tau(\boldsymbol{\eta}(\tau))-\nabla_{\boldsymbol{\eta}_1}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))| \\
&+\sup_{\tau\in[h,1-h]}|\nabla_{\boldsymbol{\eta}_1}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau))-E(\nabla_{\boldsymbol{\eta}_1}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)))-\frac{1}{T}\sum_{t=1}^T\nabla_{\boldsymbol{\vartheta}}\ell(\widetilde{\mathbf{x}}_t(\tau_t),\widetilde{\mathbf{z}}_{t-1}(\tau_t);\boldsymbol{\theta}(\tau_t))K_h(\tau_t-\tau)| \\
&+\sup_{\tau\in[h,1-h]}|E(\nabla_{\boldsymbol{\eta}_1}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)))+\frac{1}{2}h^2\widetilde{c}_2\mathbf{\Sigma}(\tau)\boldsymbol{\theta}^{(2)}(\tau)| \\
= &\sup_{\tau\in[h,1-h]}|E(\nabla_{\boldsymbol{\eta}_1}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)))+\frac{1}{2}h^2\widetilde{c}_2\mathbf{\Sigma}(\tau)\boldsymbol{\theta}^{(2)}(\tau)|+O_P((Th)^{-1}+\beta_T h^2).
\end{aligned}$$

In addition, by the proof of the first result of this lemma, we have

$$\sup_{\tau\in[h,1-h]}|E(\nabla_{\boldsymbol{\eta}_1}\widetilde{\mathscr{L}}_\tau(\boldsymbol{\eta}(\tau)))+\frac{1}{2}h^2\widetilde{c}_2\mathbf{\Sigma}(\tau)\boldsymbol{\theta}^{(2)}(\tau)|=O(h^3+(Th)^{-1}).$$

The proof is now complete. □

Proof of Lemma B.8.

(1). By the definition of class $\mathscr{H}(C,\boldsymbol{\chi},M)$ and using Hölder's inequality, we have

$$\begin{aligned}
\left\|\sup_{\boldsymbol{\vartheta}\in\mathbf{\Theta}_r}|g(\mathbf{y},\boldsymbol{\vartheta})-g(\mathbf{y}',\boldsymbol{\vartheta})|\right\|_q &\leq M\left\||\mathbf{y}-\mathbf{y}'|_{\boldsymbol{\chi}}(1+|\mathbf{y}|_{\boldsymbol{\chi}}^{C-1}+|\mathbf{y}'|_{\boldsymbol{\chi}}^{C-1})\right\|_q \\
&\leq M\,\||\mathbf{y}-\mathbf{y}'|_{\boldsymbol{\chi}}\|_{qC}\left(1+\||\mathbf{y}|_{\boldsymbol{\chi}}\|_{qC}^{C-1}+\||\mathbf{y}'|_{\boldsymbol{\chi}}\|_{qC}^{C-1}\right) \\
&\leq M\,\||\mathbf{y}-\mathbf{y}'|_{\boldsymbol{\chi}}\|_{qC}
\end{aligned}$$

provided that $\||\mathbf{y}|_{\boldsymbol{\chi}}\|_{qC}\leq\sum_{j=0}^\infty\chi_j\|\mathbf{y}_j\|_{qC}=O(1)$.
(2)–(3). Parts (2) and (3) can be proved in a similar manner as part (1).

□

Proof of Lemma B.10.

By using the summation-by-parts formula, Gaussian approximation results in Lemma B.9 and $\frac{(\log T)^4}{T^{(sq+2)/(2sq+3q-2)}h}\to 0$, we have

$$\begin{aligned}
&\sup_{\tau\in[0,1]}\left|\widetilde{\mathbf{D}}_{\widetilde{\mathbf{h}}}(\tau)-(Th)^{-1}\sum_{t=1}^T\mathbf{\Sigma}_{\widetilde{\mathbf{h}}}^{1/2}(\tau_t)\mathbf{v}_t\widehat{K}((\tau_t-\tau)/h)\right| \\
=&\sup_{\tau\in[0,1]}\left|(Th)^{-1}\widehat{K}((\tau_T-\tau)/h)\sum_{t=1}^T(\widetilde{\mathbf{h}}_t(\tau_t)-\mathbf{\Sigma}_{\widetilde{\mathbf{h}}}^{1/2}(\tau_t)\mathbf{v}_t)-\right.
\end{aligned}$$

$$
\left|(Th)^{-1}\sum_{t=1}^{T-1}(\widehat{K}((\tau_{t+1}-\tau)/h)-\widehat{K}((\tau_t-\tau)/h))\sum_{j=1}^{t}(\widetilde{\mathbf{h}}_j(\tau_j)-\mathbf{\Sigma}_{\widetilde{\mathbf{h}}}^{1/2}(\tau_j)\mathbf{v}_j)\right|
$$

$$
\leq M\cdot(Th)^{-1}\sup_{1\leq t\leq T}\left|\sum_{j=1}^{t}(\widetilde{\mathbf{h}}_j(\tau_j)-\mathbf{\Sigma}_{\widetilde{\mathbf{h}}}^{1/2}(\tau_j)\mathbf{v}_j)\right|=O_P\left((Th)^{-1}\pi_T\right)=o_P((Th\log T)^{-1/2}),
$$

where π_T is defined in Lemma B.9 and $\{\mathbf{v}_t\}$ is a sequence of i.i.d. normal vectors. In addition, by the Lipschitz property of $\mathbf{\Sigma}_{\widetilde{\mathbf{h}}}(\cdot)$, we have

$$
(Th)^{-1}\sum_{t=1}^{T}(\mathbf{\Sigma}_{\widetilde{\mathbf{h}}}^{1/2}(\tau_t)-\mathbf{\Sigma}_{\widetilde{\mathbf{h}}}^{1/2}(\tau))\mathbf{v}_t\widehat{K}((\tau_t-\tau)/h)\sim N\left(\mathbf{0},\mathbf{V}_T(\tau)\right)
$$

and $|\mathbf{V}_T(\tau)|=O(T^{-1}h)$. Hence, we have

$$
\sup_{\tau\in[0,1]}\left|(Th)^{-1}\sum_{t=1}^{T}(\mathbf{\Sigma}_{\widetilde{\mathbf{h}}}^{1/2}(\tau_t)-\mathbf{\Sigma}_{\widetilde{\mathbf{h}}}^{1/2}(\tau))\mathbf{v}_t\widehat{K}((\tau_t-\tau)/h)\right|=O_P(T^{-1/2}h^{1/2}(\log T)^{1/2})=o_P((Th\log T)^{-1/2})
$$

if $h\log T\to 0$.

Finally, by Lemma 1 in Zhou and Wu (2010), we have

$$
\lim_{T\to\infty}\Pr\left(\sqrt{\frac{Th}{\widetilde{v}_0}}\sup_{\tau\in[h,1-h]}\left|(Th)^{-1}\sum_{t=1}^{T}\mathbf{v}_t\widehat{K}((\tau_t-\tau)/h)\right|-B(m^*)\leq\frac{u}{\sqrt{2\log(m^*)}}\right)=\exp(-2\exp(-u)).
$$

Combining the above analyses, we have proved Lemma B.10. □

B.4 Computation of the Local Linear ML Estimates

In our numerical studies, we use the function *fminunc* in programming language MATLAB to minimize the negative of log-likelihood function. The initial guess is important when using optimization functions because these optimizers are trying to find a local minimum, i.e. the one closest to the initial guess that can be achieved using derivatives. In this section, we give a possible choice of initial estimates.

We could estimate the coefficients of time-varying VARMA(p,q) model

$$
\mathbf{x}_t=\sum_{j=1}^{p}\mathbf{A}_j(\tau_t)\mathbf{x}_{t-j}+\boldsymbol{\eta}_t+\sum_{j=1}^{q}\mathbf{B}_j(\tau_t)\boldsymbol{\eta}_{t-j}\quad\text{with}\quad\boldsymbol{\eta}_t=\boldsymbol{\omega}(\tau_t)\boldsymbol{\varepsilon}_t,
$$

by kernel-weighted least squares method if the lagged $\boldsymbol{\eta}_t$ were given. To obtain a preliminary estimator, we first fit a long VAR model and then use estimated residuals in place of true residuals. Consider the VAR(p_T) model

$$
\mathbf{x}_t=\sum_{j=1}^{p_T}\mathbf{\Gamma}_j(\tau_t)\mathbf{x}_{t-j}+\boldsymbol{\eta}_t,
$$

where p_T is set to be $2(Th)^{1/3}$ in our numerical studies. Then, we compute $\widehat{\boldsymbol{\eta}}_t=\mathbf{x}_t-\sum_{j=1}^{p_T}\widehat{\mathbf{\Gamma}}_j(\tau_t)\mathbf{x}_{t-j}$, where $\{\widehat{\mathbf{\Gamma}}_j(\tau)\}$ are the local linear least squares estimators. Given $\widehat{\boldsymbol{\eta}}_t$, we are able to estimate $\{\mathbf{A}_j(\tau)\}$, $\{\mathbf{B}_j(\tau)\}$ and $\mathbf{\Omega}(\tau)$ as well as their derivatives by local linear least squares method.

In order to achieve identifications, certain restrictions should be imposed on the coefficients of the VARMA model. Suppose there exists a known matrix $\mathbf{R}$ and a vector $\boldsymbol{\gamma}(\tau)$ satisfying

$$
\mathrm{vec}(\mathbf{A}_1(\tau),...,\mathbf{A}_p(\tau),\mathbf{B}_1(\tau),...,\mathbf{B}_q(\tau))=\mathbf{R}\boldsymbol{\gamma}(\tau),
$$

which follows that

$$
\mathbf{x}_t\simeq(\mathbf{z}_{t-1}^{\top}\otimes\mathbf{I}_m)\mathbf{R}[\boldsymbol{\gamma}(\tau)+\boldsymbol{\gamma}^{(1)}(\tau)(\tau_t-\tau)]+\boldsymbol{\eta}_t,
$$

where $\mathbf{z}_t=[\mathbf{x}_t^{\top},...,\mathbf{x}_{t-p+1}^{\top},\widehat{\boldsymbol{\eta}}_t^{\top},...,\widehat{\boldsymbol{\eta}}_{t-q}^{\top}]^{\top}$. Then the local linear estimator of $(\boldsymbol{\gamma}(\tau),\boldsymbol{\gamma}^{(1)}(\tau))$ is given by

$$
\begin{pmatrix}\widehat{\boldsymbol{\gamma}}(\tau)\\ h\widehat{\boldsymbol{\gamma}}^{(1)}(\tau)\end{pmatrix}=\left(\sum_{t=1}^{T}\mathbf{R}^{\top}\mathbf{Z}_{t-1}^{*}\mathbf{Z}_{t-1}^{*\top}\mathbf{R}K_h(\tau_t-\tau)\right)^{-1}\sum_{t=1}^{T}\mathbf{R}^{\top}\mathbf{Z}_{t-1}^{*}\mathbf{x}_tK_h(\tau_t-\tau),
$$

where $\mathbf{Z}_t^* = \mathbf{z}_t \otimes \mathbf{I}_m \otimes [1, \frac{\tau_{t+1}-\tau}{h}]^\top$. Similarly, the local linear estimator of $(\mathrm{vech}(\mathbf{\Omega}(\tau)), \mathrm{vech}(\mathbf{\Omega}^{(1)}(\tau)))$ is given by

$$\begin{pmatrix} \mathrm{vech}(\widehat{\mathbf{\Omega}}(\tau)) \\ h\mathrm{vech}(\widehat{\mathbf{\Omega}}^{(1)}(\tau)) \end{pmatrix} = \left(\sum_{t=1}^{T} \mathbf{Z}_t \mathbf{Z}_t^\top K_h(\tau_t - \tau)\right)^{-1} \sum_{t=1}^{T} \mathbf{Z}_t \mathrm{vech}(\widehat{\boldsymbol{\eta}}_t \widehat{\boldsymbol{\eta}}_t^\top) K_h(\tau_t - \tau),$$

where $\mathbf{Z}_t = [1, \frac{\tau_t - \tau}{h}]^\top \otimes \mathbf{I}_{m(m+1)/2}$.

We next consider the preliminary estimation of Multivariate GARCH Models. Define $\mathbf{y}_t = \mathbf{x}_t \odot \mathbf{x}_t$ and $\mathbf{v}_t = \mathbf{y}_t - \mathbf{h}_t$. We can rewrite model (1.4) as

$$\mathbf{y}_t = \mathbf{c}_0(\tau_t) + \sum_{j=1}^{\max(p,q)} (\mathbf{C}_j(\tau_t) + \mathbf{D}_j(\tau_t))\mathbf{y}_{t-j} + \mathbf{v}_t + \sum_{j=1}^{q} (-\mathbf{D}_j(\tau_t))\mathbf{v}_{t-j}$$

with $E(\mathbf{v}_t | \mathscr{F}_{t-1}) = 0$. Similar to the VARMA model, we are able to estimate $\mathbf{c}_0(\tau)$, $\{\mathbf{C}_j(\tau)\}$ and $\{\mathbf{D}_j(\tau)\}$ as well as their derivatives by local linear least squares method. Consider the VAR(p_T) model

$$\mathbf{y}_t = \sum_{j=1}^{p_T} \mathbf{\Phi}_j(\tau_t) \mathbf{y}_{t-j} + \mathbf{v}_t,$$

where p_T is set to be $2(Th)^{1/3}$ in our numerical studies. Then, we compute $\widehat{\mathbf{v}}_t = \mathbf{y}_t - \sum_{j=1}^{p_T} \widehat{\mathbf{\Phi}}_j(\tau_t)\mathbf{p}_{t-j}$, $\widehat{\mathbf{h}}_t = \mathbf{y}_t - \widehat{\mathbf{v}}_t$ and $\widehat{\boldsymbol{\eta}}_t = \mathrm{diag}^{-1/2}(\widehat{\mathbf{h}}_t)\mathbf{x}_t$. Hence, the local linear estimator of $(\mathrm{vechl}(\mathbf{\Omega}(\tau)), \mathrm{vechl}(\mathbf{\Omega}^{(1)}(\tau)))$ is given by

$$\begin{pmatrix} \mathrm{vechl}(\widehat{\mathbf{\Omega}}(\tau)) \\ h\mathrm{vechl}(\widehat{\mathbf{\Omega}}^{(1)}(\tau)) \end{pmatrix} = \left(\sum_{t=1}^{T} \mathbf{Z}_t \mathbf{Z}_t^\top K_h(\tau_t - \tau)\right)^{-1} \sum_{t=1}^{T} \mathbf{Z}_t \mathrm{vechl}(\widehat{\boldsymbol{\eta}}_t \widehat{\boldsymbol{\eta}}_t^\top) K_h(\tau_t - \tau),$$

where $\mathbf{Z}_t = [1, \frac{\tau_t - \tau}{h}]^\top \otimes \mathbf{I}_{m(m-1)/2}$ and $\mathrm{vechl}(\cdot)$ stacks the lower triangular part of a square matrix excluding the diagonal.

Finally, we consider the preliminary estimation of time-varying VARMA-GARCH models. In this case, we first estimate the VARMA part and then use estimated residuals to estimate the GARCH part. Consider the VAR(p_T) model with GARCH-type errors

$$\mathbf{x}_t = \sum_{j=1}^{p_T} \mathbf{\Gamma}_j(\tau_t) \mathbf{x}_{t-j} + \mathbf{v}_t,$$

where p_T is set to be $2(Th)^{1/3}$ in our numerical studies. Then, we compute $\widehat{\mathbf{v}}_t = \mathbf{x}_t - \sum_{j=1}^{p_T} \widehat{\mathbf{\Gamma}}_j(\tau_t)\mathbf{x}_{t-j}$, where $\{\widehat{\mathbf{\Gamma}}_j(\tau)\}$ are the local linear least squares estimators. Given $\widehat{\boldsymbol{\eta}}_t$, we are able to estimate $\{\mathbf{A}_j(\tau)\}$ and $\{\mathbf{B}_j(\tau)\}$ as well as their derivatives by local linear least squares method as stated above. In addition, based on estimated residuals $\widehat{\mathbf{v}}_t$, we are able to estimate the GARCH part in a similar manner as above.

B.5 Additional Simulation Results

In this appendix we report some additional simulation results for time-varying GARCH models. The data generating process is specified as follows:

$$\text{DGP 3}: \mathbf{x}_t = \mathrm{diag}(h_{1,t}^{1/2}, \ldots, h_{m,t}^{1/2})\boldsymbol{\eta}_t,$$

where $\boldsymbol{\eta}_t = \mathbf{\Omega}^{1/2}(\tau_t)\boldsymbol{\varepsilon}_t$, $\mathbf{h}_t = \mathbf{c}_0(\tau_t) + \mathbf{C}_1(\tau_t)(\mathbf{x}_{t-1} \odot \mathbf{x}_{t-1}) + \mathbf{D}_1(\tau_t)\mathbf{h}_{t-1}$, $\{\boldsymbol{\varepsilon}_t\}$ are i.i.d. draws from $N(\mathbf{0}_{2\times 1}, \mathbf{I}_2)$, $\mathbf{c}_0(\tau) = [2\exp\{0.5\tau - 0.5\}, 3 + 0.2\cos(\tau)]^\top$,

$$\mathbf{C}_1(\tau) = \begin{bmatrix} 0.4 + 0.05\cos(\tau) & 0.05(\tau - 0.5)^2 \\ 0.05(\tau - 0.5)^2 & 0.4 + 0.05\sin(\tau) \end{bmatrix},$$
$$\mathbf{D}_1(\tau) = \begin{bmatrix} 0.4 - 0.1\cos(\tau) & 0 \\ 0 & 0.3 - 0.1\sin(\tau) \end{bmatrix},$$
$$\mathbf{\Omega}(\tau) = \begin{bmatrix} 1 & 0.3\sin(\tau) \\ 0.3\sin(\tau) & 1 \end{bmatrix}.$$

We set the order of GARCH process to be $(1,1)$ since GARCH$(1,1)$ models are typically used in practice and higher order GARCH models are unnecessary (cf., Andreou and Werker, 2015).

We present the empirical coverage probabilities associated with the UCB in Table B.1. Again, we find that the conditional variance model requires more data than the conditional mean model to achieve a reasonable finite sample performance.

Table B.1: Empirical Coverage Probabilities of the UCB for DGP 3

	$\widetilde{h}$	$\mathbf{c}_0(\cdot)$	$\mathbf{C}_1(\cdot)$	$\mathbf{D}_1(\cdot)$	$\mathbf{\Omega}(\cdot)$
$T = 1000$	0.55	0.869	0.876	0.838	0.945
	0.60	0.882	0.866	0.843	0.945
	0.65	0.889	0.872	0.859	0.945
	0.70	0.892	0.881	0.871	0.950
$T = 2000$	0.50	0.897	0.881	0.901	0.950
	0.55	0.892	0.881	0.903	0.940
	0.60	0.900	0.888	0.910	0.950
	0.65	0.907	0.889	0.910	0.950
$T = 4000$	0.35	0.929	0.932	0.943	0.920
	0.4	0.950	0.944	0.943	0.919
	0.45	0.950	0.947	0.946	0.950
	0.50	0.929	0.944	0.946	0.960

References

Hall, P. and Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, Academic Press.

Jeantheau, T. (1998), 'Strong consistency of estimators for multivariate ARCH models', *Econometric theory* **14**(1), 70–86.

Wu, W. B. and Zhou, Z. (2011), 'Gaussian approximations for non-stationary multiple time series', *Statistica Sinica* pp. 1397–1413.

Zhang, D. and Wu, W. B. (2017), 'Gaussian approximation for high dimensional time series', *Annals of Statistics* **45**(5), 1895–1919.

Zhou, Z. and Wu, W. B. (2010), 'Simultaneous inference of linear models with time varying coefficients', *Journal of the Royal Statistical Society: Series B* **72**(4), 513–531.