MONASH University

MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

Asymptotics for Time-Varying Vector MA (∞) Processes

Yayi Yan, Jiti Gao and Bin Peng

December 2021

Working Paper 22/21

AACSB ACCREDITED EQUIS ACCREDITED AMBA ACCREDITED

Asymptotics for Time–Varying Vector MA(∞) Processes

Yayi Yan, Jiti Gao and Bin Peng

Monash University

December 6, 2021

Abstract

Moving average infinity (MA(∞)) processes play an important role in modeling time series data. While a strand of literature on time series analysis emphasizes the importance of modeling smooth changes over time and therefore is shifting its focus from parametric models to nonparametric ones, MA(∞) processes with constant parameters are often part of the fundamental data generating mechanism. Along this line of research, an intuitive question is how to allow the underlying data generating mechanism evolves over time. To better capture the dynamics, this paper considers a new class of time–varying vector moving average infinity (VMA(∞)) processes. Accordingly, we establish some new asymptotic properties, including the law of large numbers, the uniform convergence, the central limit theory, the bootstrap consistency, and the long–run covariance matrix estimation for the class of time–varying VMA(∞) processes. Finally, we demonstrate the empirical relevance and usefulness of the newly proposed model and estimation theory through extensive simulated and real data studies.

Keywords: Multivariate Time Series; Nonparametric Kernel Estimation; Time–Varying Beveridge–Nelson Decomposition

JEL Classification: C14, C32, E52

Corresponding author: Jiti Gao, Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria 3145, Australia. Email: `Jiti.Gao@monash.edu`.

1 Introduction

Moving average infinity (MA(∞)) processes are possibly one of the most fundamental data generating mechanisms when studying time series (Beveridge and Nelson, 1981; Phillips and Solo, 1992; Hamilton, 1994; Lütkepohl, 2005). For example, in the field of macroeconomics, MA(∞) representations of multivariate time series with time–invariant coefficients play a central role when estimating impulse response functions, which trace the transmission mechanism of economic shocks and are useful for policy analyses. Such a routine exercise has also been widely adopted in the fields such as signal processing, climatology (e.g., Bühlmann, 1998; Friedrich et al., 2020; Paul, 2020; Plagborg-Møller and Wolf, 2021). As pointed out by Hansen (2001), it seems that dynamic models with time–invariant coefficients may be unnecessarily restrictive in order to accommodate smooth changes over a period of time rather than in a static manner. To model such time–varying behaviours, an important strand of the relevant literature assumes that the coefficients of dynamic models evolve in a random way (e.g., Primiceri, 2005; Petrova, 2019), and estimation procedures heavily depend on Bayesian computational algorithms.

Compared with the aforementioned literature, nonparametric methods have also been proposed to estimate deterministically unknown time–varying parameters involved in specific autoregressive models. Up to this point, it is worth bringing up the terminology "local stationarity", which at least dates back to the seminal work by Dahlhaus (1996). There have since been some other studies on univariate autoregressive regression models (Dahlhaus and Rao, 2006; Dahlhaus, 2012; Zhang and Wu, 2012; Richter and Dahlhaus, 2019) in recent years. To the best of our knowledge, it has had very little success to generalize the local stationarity approach to multivariate autoregressive settings. In specific cases where different locally stationary univariate time series may be approximated by their stationary versions on the same segments, the local stationarity technique may be applicable to such specific multivariate settings. However, univariate time series components of general multivariate time series may have quite different behaviours, and therefore different locally stationary univariate time series may not be approximated by their stationary versions on the same segments, such as the three univariate time series plotted in Figure 1 of Section 5.3.

In order to address estimation and inferential issues for general multivariate dynamic models, we show the versatility of an alternative approach that is designed for a wide class of time–varying vector moving average (VMA)(∞) processes associated with nonparametrically unknown time–varying coefficients. In particular, we develop an explicit decomposition for partial sums of time–varying VMA(∞) processes into the long–run and transitory elements, which is known as the Beveridge–Nelson (BN) decomposition (Beveridge and Nelson, 1981; Phillips and Solo, 1992). The long–run component of the decomposition yields a martingale approximation, which ensures the feasibility of achieving a variety of asymptotic properties for multivariate dynamic models. The proposed time–varying BN decomposition then facilitates the establishment of a number of new asymptotic properties for the proposed estimators of the unknown trend functions and coefficient matrices of the general class of VMA(∞) models under some very mild assumptions. We also show that the VMA(∞) process naturally covers a class of time–varying VARX models, and further establish several asymptotic properties of non- and semi-parametric estimators in (partially) time–varying VARX models. In the empirical study, we apply the newly proposed framework to study the long–run level of inflation and the natural rate of unemployment. We find that (i) the long–run level of inflation is more anchored now and is close to the Federal Reserve's target of two percent after the beginning of the Great Moderation period, and (ii) the natural rate of unemployment is less persistent and increases rapidly during "Second Oil Crisis" and "Global Financial Crisis".

In summary, our contributions are as follows. First, we propose a new class of time–varying VMA(∞) models, and then develop a time–varying counterpart of the conventional BN decomposition before we are able to establish a variety of asymptotic properties for the estimation of the unknown trends and coefficients of the general class of time–varying VMA(∞) models, such as the law of large numbers, the uniform convergence and the central limit theory, under some very mild assumptions. Second, we propose a dependent wild bootstrap (DWB) procedure and a heteroscedasticity and autocorrelation consistent (HAC) covariance matrix estimation method to ensure that the proposed estimation theory is valid for inferential purposes and empirical implementations, and the finite–sample evaluation results show that the proposed estimation and inferential methods work well numerically. Third, after employing the time–varying BN decomposition technology, we are able to consistently estimate time–varying coefficients involved in a class of VARX models using non– and semi–parametric kernel methods.

The paper is organised as follows. Section 2 introduces a class of time–varying VMA(∞) processes, develops a time–varying counterpart of the conventional BN decomposition, and establishes a set of asymptotic properties. Section 3 applies the results of Section 2 to establish an inferential theory for smooth deterministic trends of the time–varying VMA(∞) model. Section 4 establishes an estimation theory for the time–varying coefficients involved in a class of time–varying VARX models. Section 5 evaluates the finite sample performance of the proposed methods through extensive simulated and real data studies. Section 6 gives a short conclusion. The preliminary lemmas and the proofs of the main results are given in Appendix A, while the proofs of the preliminary lemmas are given in Appendix B.

Before proceeding further, it is convenient to introduce some notation: $\|\cdot\|$ denotes the Euclidean norm of a vector or the Frobenius norm of a matrix; $\otimes$ denotes the Kronecker product; $\boldsymbol{I}_a$ stands for an $a\times a$ identity matrix; $\mathbf{0}_{a\times b}$ stands for an $a\times b$ matrix of zeros, and we write $\mathbf{0}_a$ for short when $a=b$; for a function $g(w)$, let $g^{(j)}(w)$ be the j^{th} derivative of $g(w)$, where $j\geq 0$ and $g^{(0)}(w)\equiv g(w)$; $K_h(\cdot)=K(\cdot/h)/h$, where $K(\cdot)$ and h stand for a nonparametric kernel function and a bandwidth respectively; let $\tilde{c}_k=\int_{-1}^{1}u^kK(u)du$ and $\tilde{v}_k=\int_{-1}^{1}u^kK^2(u)du$ for integer $k\geq 0$; vec$(\cdot)$ stacks the elements of an $m\times n$ matrix as an $mn\times 1$ vector; $\mathrm{tr}\,(\boldsymbol{A})$ denotes the trace of $\boldsymbol{A}$; finally, let $\to_P$ and $\to_D$ denote convergence in probability and convergence in distribution, respectively.

2 The Setup with Asymptotics

Consider the following time–varying VMA(∞) model:

$$\boldsymbol{x}_t = \boldsymbol{\mu}_t + \sum_{j=0}^{\infty} \boldsymbol{B}_{j,t}\boldsymbol{\epsilon}_{t-j} := \boldsymbol{\mu}_t + \mathbb{B}_t(L)\boldsymbol{\epsilon}_t, \quad t = 1, \ldots, T, \tag{2.1}$$

where $\boldsymbol{x}_t$ is a vector of d–dimensional observable variables, $\boldsymbol{\mu}_t$ is a vector of d–dimensional unknown deterministic trending functions, $\boldsymbol{B}_{j,t}$'s are $d \times d$ unknown matrices, $\boldsymbol{\epsilon}_t$ is a vector of d–dimensional random innovations, and d is fixed. Obviously, $\mathbb{B}_t(L) = \sum_{j=0}^{\infty} \boldsymbol{B}_{j,t} L^j$, where L is the lag operator. Throughout this paper, we impose the following necessary conditions to establish our asymptotic properties.

Assumption 1. $\max_{t\geq 1} \sum_{j=1}^{\infty} j\|\boldsymbol{B}_{j,t}\| < \infty$, $\limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{j=1}^{\infty} j\|\boldsymbol{B}_{j,t+1} - \boldsymbol{B}_{j,t}\| < \infty$, *and* $\limsup_{T\to\infty} \sum_{t=1}^{T-1} \|\boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}_t\| < \infty$.

Assumption 2. $\{\boldsymbol{\epsilon}_t\}_{t=-\infty}^{\infty}$ *is a martingale difference sequences (m.d.s.) adapted to the filtration* $\{\mathcal{F}_t\}$, *where* $\mathcal{F}_t = \sigma(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{t-1}, \ldots)$ *is the* σ*–field generated by* $(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{t-1}, \ldots)$, $E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top|\mathcal{F}_{t-1}] = \boldsymbol{I}_d$ *almost surely (a.s.), and* $\max_{t\geq 1} E\|\boldsymbol{\epsilon}_t\|^{\delta} < \infty$ *for some* $\delta \geq 4$.

Assumption 1 regulates the matrices $\boldsymbol{B}_{j,t}$'s, and ensures the validity of the BN decomposition under a time–varying framework. It covers many cases, including (i) the parametric setting of Phillips and Solo (1992), and (ii) $\boldsymbol{B}_{j,t} := \boldsymbol{B}_j(\tau_t)$, where $\tau_t := t/T$ and $\boldsymbol{B}_j(\cdot)$ satisfies Lipschitz continuity on $[0, 1]$ for all j. These conditions can be easily verified as they are directly related to some commonly used data generating mechanisms, see, for example, Proposition 2.1 below. Assumption 2 imposes conditions on the innovation error terms by replacing the commonly used independent and identically distributed (i.i.d.) innovations (e.g., Dahlhaus and Polonik, 2009) with a martingale difference structure.

We are now ready to comment on the usefulness of (2.1). An application of the BN decomposition to (2.1) immediately yields:

$$\boldsymbol{x}_t = \boldsymbol{\mu}_t + \mathbb{B}_t(1)\boldsymbol{\epsilon}_t + \widetilde{\mathbb{B}}_t(L)\boldsymbol{\epsilon}_{t-1} - \widetilde{\mathbb{B}}_t(L)\boldsymbol{\epsilon}_t, \tag{2.2}$$

where $\mathbb{B}_t(L) = \mathbb{B}_t(1) - (1-L)\widetilde{\mathbb{B}}_t(L)$, $\widetilde{\mathbb{B}}_t(L) = \sum_{j=0}^{\infty} \widetilde{\boldsymbol{B}}_{j,t} L^j$, and $\widetilde{\boldsymbol{B}}_{j,t} = \sum_{k=j+1}^{\infty} \boldsymbol{B}_{k,t}$. The decomposition of (2.2) allows one to derive asymptotic properties associated with $\boldsymbol{x}_t$'s. For example, the following lemma holds under Assumptions 1 and 2.

Lemma 2.1. *Under Assumptions 1–2, as* $T \to \infty$, *for* $\forall r \in [0, 1]$

$$\frac{1}{\sqrt{T}} \boldsymbol{\Sigma}^{-1/2}(r) \sum_{t=1}^{\lfloor Tr \rfloor} (\boldsymbol{x}_t - \boldsymbol{\mu}_t) \to_D \boldsymbol{W}(r),$$

where $\boldsymbol{\Sigma}(r) := \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{\lfloor Tr \rfloor} \mathbb{B}_t(1)\mathbb{B}_t^\top(1)$ *and* $\boldsymbol{W}(\cdot)$ *is a standard multivariate Brownian motion.*

By Lemma 2.1, it is easy to see that (2.1) extends similar treatments by Phillips and Solo (1992) for the univariate linear process case, and allows one to relax many $I(0)$ and $I(1)$ related results of the literature using a time–varying VMA(∞) framework.

Below, we list several examples, of which the parametric counterparts can be found in classic textbooks (e.g., Lütkepohl, 2005).

Example 1. Suppose that $\boldsymbol{x}_t$ is a d–dimensional time–varying VAR(p) process:

$$\boldsymbol{x}_t = \boldsymbol{A}_{1,t}\boldsymbol{x}_{t-1} + \cdots + \boldsymbol{A}_{p,t}\boldsymbol{x}_{t-p} + \boldsymbol{\epsilon}_t, \tag{2.3}$$

which has been widely studied in the literature with Bayesian framework being the dominant approach (e.g., Benati and Surico, 2009; Paul, 2020). Similar to Hamilton (1994, p.260), model (2.3) can be expressed as a time–varying VMA(∞) process $\boldsymbol{x}_t = \sum_{j=0}^{\infty} \boldsymbol{B}_{j,t}\boldsymbol{\epsilon}_{t-j}$, where $\boldsymbol{B}_{0,t} = \boldsymbol{I}_d$, $\boldsymbol{B}_{j,t} = \boldsymbol{J}\prod_{i=0}^{j-1}\boldsymbol{\Phi}_{t-i}\boldsymbol{J}^\top$ for $j \geq 1$, $\boldsymbol{J} = [\boldsymbol{I}_d, \boldsymbol{0}_{d\times d(p-1)}]$ and $\boldsymbol{\Phi}_t$ is the companion matrix.

Example 2. Suppose that $\boldsymbol{x}_t$ is a d–dimensional time–varying VARMA(p, q) process as follows:

$$\boldsymbol{x}_t = \boldsymbol{A}_{1,t}\boldsymbol{x}_{t-1} + \cdots + \boldsymbol{A}_{p,t}\boldsymbol{x}_{t-p} + \boldsymbol{\epsilon}_t + \boldsymbol{\Theta}_{1,t}\boldsymbol{\epsilon}_{t-1} + \cdots + \boldsymbol{\Theta}_{q,t}\boldsymbol{\epsilon}_{t-q}. \tag{2.4}$$

Simple algebra shows that model (2.4) can be expressed as $\boldsymbol{x}_t = \sum_{b=0}^{\infty} \boldsymbol{D}_{b,t}\boldsymbol{\epsilon}_{t-b}$ with $\boldsymbol{D}_{b,t} = \sum_{j=\max(0,b-q)}^{b} \boldsymbol{B}_{j,t}\boldsymbol{\Theta}_{b-j,t-j}$, in which $\boldsymbol{B}_{j,t}$ is defined similarly as in Example 1, and $\boldsymbol{\Theta}_{0,t} = \boldsymbol{I}_d$ is independent of t.

Example 3. Suppose that $\boldsymbol{x}_t$ is a d–dimensional time–varying VARX process of the form:

$$\boldsymbol{x}_t = \boldsymbol{A}_{1,t}\boldsymbol{x}_{t-1} + ... + \boldsymbol{A}_{p,t}\boldsymbol{x}_{t-p} + \boldsymbol{\Theta}_t\boldsymbol{z}_t + \boldsymbol{\epsilon}_t \quad \text{and} \quad \boldsymbol{z}_t = \sum_{j=0}^{\infty} \boldsymbol{C}_{j,t}\boldsymbol{v}_{t-j}, \tag{2.5}$$

where $\boldsymbol{z}_t$ is an m–dimensional vector and $\boldsymbol{\Theta}_t$ is a $d \times m$ matrix. Then model (2.5) can be further written as

$$\begin{bmatrix} \boldsymbol{x}_t \\ \boldsymbol{z}_t \end{bmatrix} = \sum_{j=0}^{\infty} \begin{bmatrix} \boldsymbol{B}_{j,t} & \boldsymbol{D}_{j,t} \\ \boldsymbol{0} & \boldsymbol{C}_{j,t} \end{bmatrix} \begin{bmatrix} \boldsymbol{\epsilon}_{t-j} \\ \boldsymbol{v}_{t-j} \end{bmatrix},$$

where $\boldsymbol{D}_{j,t} = \sum_{k=0}^{j} \boldsymbol{B}_{k,t}\boldsymbol{\Theta}_{t-k}\boldsymbol{C}_{j-k,t-k}$ and $\boldsymbol{B}_{j,t}$ is defined in a way similar to that in Example 1.

The following proposition shows that each of the models listed in the above three examples may have a time–varying VMA(∞) representation.

Proposition 2.1.

1. *Consider Examples 1 and 2. Suppose that the roots of* $\boldsymbol{I}_d - \boldsymbol{A}_{1,t} - \cdots - \boldsymbol{A}_{p,t} = \boldsymbol{0}_d$ *all lie outside the unit circle uniformly over* t, $\limsup_{T\to\infty}\sum_{t=1}^{T-1}\|\boldsymbol{A}_{m,t+1} - \boldsymbol{A}_{m,t}\| < \infty$ *for* $m = 1, \ldots, p$ *and* $\boldsymbol{A}_{m,t} = \boldsymbol{A}_{m,1}$ *for* $t \leq 0$ *and* $m = 1, \ldots, p$. *In addition, suppose that in Example 2,*

$\limsup_{T\to\infty}\sum_{t=1}^{T-1}\|\mathbf{\Theta}_{m,t+1}-\mathbf{\Theta}_{m,t}\|<\infty$ *for* $m=1,\ldots,q$. *Then both* (2.3) *and* (2.4) *are time–varying VMA*(∞) *processes, for which the coefficients satisfy Assumption 1.*

2. *Suppose that* $\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty} j\|\boldsymbol{C}_{j,t+1}-\boldsymbol{C}_{j,t}\|<\infty$ *and* $\max_{t\geq 1}\sum_{j=1}^{\infty} j\|\boldsymbol{C}_{j,t}\|<\infty$ *for Example 3. Moreover, let* $\boldsymbol{A}_{j,t}$ *and* $\mathbf{\Theta}_t$ *satisfy the same conditions as those in the first result of this proposition. Then* (2.5) *is a time–varying VMA*(∞)*, for which the coefficients satisfy Assumption 1.*

Estimation and testing issues for the coefficient matrices involved in Examples 1–3 and their semiparametric counterparts are of significant interest in econometrics, and should be fully investigated in separate research. Section 3 of the original working paper version of this paper by Yan et al. (2020) has discussed some estimation problems for Example 2.1. Section 4 below considers non– and semi–parametric estimation problems for Example 3.

2.1 Asymptotic Properties of the Sample Moments

In this subsection, we present some useful asymptotic properties associated with (2.1). First, we propose the law of large numbers for two weighted sample moments of $\boldsymbol{x}_t$.

Lemma 2.2. *Let Assumptions 1 and 2 hold. In addition, suppose that* $\{\boldsymbol{W}_{T,t}\}_{t=1}^{T}$ *is a sequence of* $m\times d$ *deterministic weighting matrices satisfying (1)* $\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}\|=O(1)$*, (2)* $\max_{t\geq 1}\|\boldsymbol{W}_{T,t}\|=O(d_T)$*, and (3)* $\sum_{t=1}^{T-1}\|\boldsymbol{W}_{T,t+1}-\boldsymbol{W}_{T,t}\|=O(d_T)$*, where* $d_T=\max_{t\geq 1}\|\boldsymbol{W}_{T,t}\|\to 0$*. Then, as* $T\to\infty$*,*

$$\sum_{t=1}^{T}\boldsymbol{W}_{T,t}\left(\boldsymbol{x}_t-E(\boldsymbol{x}_t)\right)=O_P(\sqrt{d_T})\quad\textit{and}\quad\sum_{t=1}^{T}\boldsymbol{W}_{T,t}\left(\boldsymbol{x}_t\boldsymbol{x}_{t+p}^{\top}-E\left(\boldsymbol{x}_t\boldsymbol{x}_{t+p}^{\top}\right)\right)=O_P(\sqrt{d_T}),$$

where both m (≥ 1) *and* p (≥ 0) *are fixed integers.*

Lemma 2.2 provides the conditions that ensure the convergence of weighted sample moments, and can be easily applied to study weighted-least-squares type of estimators. It is worth stressing that the conditions on $\{\boldsymbol{W}_{T,t}\}$ are weak, in the sense that Lemma 2.2 covers both the parametric rate $d_T=\frac{1}{T}$ (e.g., $\boldsymbol{W}_{T,t}=\frac{1}{T}$) and the nonparametric rate $d_T=\frac{1}{Th}$ (e.g., $\boldsymbol{W}_{T,t}=\frac{1}{T}K_h(\tau_t-\tau)$ for $\forall\tau\in[0,1]$).

Next, we strengthen the results of Lemma 2.2, and establish the rates of uniform convergence.

Lemma 2.3. *Let Assumptions 1 and 2 hold. In addition, let* $\{\boldsymbol{W}_{T,t}(\cdot)\}_{t=1}^{T}$ *be a sequence of* $m\times d$ *matrices of deterministic weighting functions, in which* m *is fixed, and each functional component is Lipschitz continuous and defined on a compact set* $[a,b]$*. Moreover, suppose that*

1. $\sup_{\tau\in[a,b]}\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}(\tau)\|=O(1)$*;*

2. $\sup_{\tau\in[a,b]}\sum_{t=1}^{T-1}\|\boldsymbol{W}_{T,t+1}(\tau)-\boldsymbol{W}_{T,t}(\tau)\|=O(d_T)$*, where* $d_T=\sup_{\tau\in[a,b],t\geq 1}\|\boldsymbol{W}_{T,t}(\tau)\|\to 0$*.*

Then as $T \to \infty$,

1. $\sup_{\tau\in[a,b]} \| \sum_{t=1}^{T} \boldsymbol{W}_{T,t}(\tau)\,(\boldsymbol{x}_t - E(\boldsymbol{x}_t))\,\| = O_P(\sqrt{d_T \log T})$ *provided* $T^{\frac{2}{\delta}} d_T \log T \to 0$*;*

2. $\sup_{\tau\in[a,b]} \| \sum_{t=1}^{T} \boldsymbol{W}_{T,t}(\tau)\left(\boldsymbol{x}_t\boldsymbol{x}_{t+p}^\top - E(\boldsymbol{x}_t\boldsymbol{x}_{t+p}^\top)\right)\| = O_P(\sqrt{d_T \log T})$ *for any fixed integer* $p \geq 0$ *provided* $T^{\frac{4}{\delta}} d_T \log T \to 0$ *and* $\max_{t\geq 1} E(\|\boldsymbol{\epsilon}_t\|^4|\mathcal{F}_{t-1}) < \infty$ *a.s., where* $\delta > 4$ *is the same as that in Assumption 2.*

Lemma 2.3 corresponds to some existing uniform convergence results for nonparametric estimation of a number of time series models associated with either stationarity or unit roots, such as those in Hansen (2008), Gao et al. (2015), Li et al. (2016), and Phillips et al. (2017). As a specific application, in Section 4 we apply this result to establish an estimation theory for a class of nonparametric and semiparametric time–varying VARX models.

2.2 Inferences

To obtain valid inferences in practice, in what follows we establish a central limit theory in Lemma 2.4, and then propose two methods (i.e., the dependent wild bootstrap (DWB) approach and the heteroscedasticity and autocorrelation consistent (HAC) covariance matrix estimation approach) to estimate an asymptotic covariance matrix in Lemmas 2.5 and 2.6, respectively.

Lemma 2.4. *Let Assumptions 1–2 hold. Suppose* $\{\boldsymbol{W}_{T,t}\}_{t=1}^{T}$ *is a sequence* $m \times d$ *deterministic weighting matrices satisfying (1)* $\sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}\| = O(1)$*, (2)* $\max_{t\geq 1} \|\boldsymbol{W}_{T,t}\| = O(d_T)$ *and (3)* $\sum_{t=1}^{T-1} \|\boldsymbol{W}_{T,t+1} - \boldsymbol{W}_{T,t}\| = O(d_T)$*, where the sequence of real numbers,* d_T*, is chosen to ensure that* $\boldsymbol{\Sigma}_{\boldsymbol{W}} = \lim_{T\to\infty} \frac{1}{d_T} \sum_{t=1}^{T} \boldsymbol{W}_{T,t}\mathbb{B}_t(1)\mathbb{B}_t^\top(1)\boldsymbol{W}_{T,t}^\top$ *is a positive definite matrix. As* $T \to \infty$*, we then have*

$$\frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \boldsymbol{W}_{T,t}\,(\boldsymbol{x}_t - E(\boldsymbol{x}_t)) \to_D N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{W}}),$$

where m (≥ 1) *is a fixed positive integer.*

With Lemma 2.4 in hand, in order to infer the population mean of $\boldsymbol{x}_t$, the only missing piece is the information of $\boldsymbol{\Sigma}_{\boldsymbol{W}}$, which is a type of long–run covariance matrix arising from the infinity memory of $\boldsymbol{x}_t$. To recover $\boldsymbol{\Sigma}_{\boldsymbol{W}}$, we then consider two approaches: (i) the DWB approach, and (ii) the HAC covariance matrix estimation. Both approaches date back to Shao (2010), and Newey and West (1987), respectively.

We start with the DWB method, and suppose that $\{\xi_t^*\}_{t=1}^{T}$ is a sequence of l–dependent time series satisfying $E[\xi_t^*] = 0$, $E[\xi_t^{*2}] = 1$, $E|\xi_t^*|^\delta < \infty$ for some $\delta > 2$, and $E[\xi_t^*\xi_s^*] = a\left((t-s)/l\right)$ for a kernel function $a(\cdot)$ and a tuning parameter l. The DWB procedure requires a tuning parameter l, which is the "block length" (Shao, 2010) ensuing the variables further than l units apart are independent.

Lemma 2.5. *Let $l \to \infty$ and $l\sqrt{d_T} \to 0$. Additionally, let $a(\cdot)$ be a symmetric kernel and Lipschitz continuous on $[-1, 1]$ satisfying that $a(0) = 1$ and*

$$K_a(x) = \int_{-\infty}^{\infty} a(u)e^{-iux}\mathrm{d}u \geq 0 \quad \textit{for } x \in \mathbb{R}.$$

Under the conditions of Lemma 2.4, as $T \to \infty$,

$$\sup_{\boldsymbol{w}\in\mathbb{R}^d} \left| \Pr^* \left[\frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \widetilde{\boldsymbol{x}}_t \xi_t^* \leq \boldsymbol{w} \right] - \Pr \left[\frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \widetilde{\boldsymbol{x}}_t \leq \boldsymbol{w} \right] \right| = o_P(1),$$

where $\widetilde{\boldsymbol{x}}_t = \boldsymbol{W}_{T,t} (\boldsymbol{x}_t - E(\boldsymbol{x}_t))$, and $\Pr^*$ denotes the probability measure induced by the DWB procedure.

Lemma 2.5 establishes the consistency of the bootstrap version of the weighted sample mean of $\boldsymbol{x}_t$. The condition of $K_a(x)$ ensures the semi–positive definiteness of the covariance matrix of $\{\xi_t^*\}_{t=1}^T$, while the restrictions on $a(\cdot)$ are satisfied by a few commonly used kernels, such as the Bartlett and Parzen kernels.

We now consider the HAC approach for us to deal with inferential issues. Specifically, we define

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{W}} = \widehat{\boldsymbol{\Xi}}_0 + \sum_{i=1}^{b} \psi\left(i/b\right) (\widehat{\boldsymbol{\Xi}}_i + \widehat{\boldsymbol{\Xi}}_i^\top), \tag{2.6}$$

where $\widehat{\boldsymbol{\Xi}}_i(\tau) = \frac{1}{d_T} \sum_{t=1}^{T-i} \boldsymbol{W}_{T,t}\boldsymbol{e}_t\boldsymbol{e}_{t+i}^\top \boldsymbol{W}_{T,t+i}^\top$ for $i \geq 0$, $\boldsymbol{e}_t = \boldsymbol{x}_t - E(\boldsymbol{x}_t)$, $\psi(\cdot)$ is a kernel function, and b is the bandwidth diverging at a relatively slow rate, in which $E(\boldsymbol{x}_t)$ is assumed to be computable at this stage. Otherwise, it will be replaced by an estimated version as in equation (3.3) below. Under some mild conditions, we establish asymptotic properties for (2.6) in the following lemma.

Lemma 2.6. *Suppose that $\psi(\cdot)$ is Lipschitz continuous, and has a compact support on $[-1, 1]$ with $\psi(0) = 1$. Additionally, let $b \to \infty$ and $b\sqrt{d_T} \to 0$. Under the conditions of Theorem 2.4, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{W}} = \boldsymbol{\Sigma}_{\boldsymbol{W}} + o_P(1)$.*

The conditions on b and $\psi(\cdot)$ are standard, and are similar to those for the DWB method. For the case of parametric estimation ($d_T = \frac{1}{T}$), the condition $b\sqrt{d_T} \to 0$ is identical to that of Hansen (1992), who proves the consistency of long–run covariance matrix estimator under the mixing condition. Apart from constructing confidence intervals for weighted sample mean, the long–run covariance estimation is also essential in model specification testing, see, for example, Zhang and Wu (2011).

Up to this point, we have established a set of asymptotic properties for the VMA(∞) process of (2.1). In the following section, we apply these results to study the smooth deterministic trends of (2.1) for the purpose of estimation and inference.

3 On the Deterministic Trends

To facilitate the development, it is useful to impose the following specifications:

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}(\tau_t) \quad \text{and} \quad \boldsymbol{B}_{j,t} = \boldsymbol{B}_j(\tau_t),$$

where $\tau_t = t/T$. Thus, (2.1) can be rewritten as

$$\boldsymbol{x}_t = \boldsymbol{\mu}(\tau_t) + \sum_{j=0}^{\infty} \boldsymbol{B}_j(\tau_t)\boldsymbol{\epsilon}_{t-j}. \tag{3.1}$$

Below we show that the trending function $\boldsymbol{\mu}(\tau)$ and the long–run covariance matrix associated with (3.1) can be well recovered, although we are unable to consistently estimate each individual $\boldsymbol{B}_j(\tau)$. The following assumptions are necessary.

Assumption 3. *Each component of $\boldsymbol{\mu}(\cdot)$ and $\boldsymbol{B}_j(\cdot)$'s is second order continuously differentiable on $[0,1]$. Moreover, $\sup_{\tau\in[0,1]} \sum_{j=1}^{\infty} j\|\boldsymbol{B}_j^{(\ell)}(\tau)\| < \infty$ for $\ell = 0, 1$.*

Assumption 4. *Let $K(\cdot)$ be a symmetric and positive kernel function defined on $[-1,1]$ with $\int_{-1}^{1} K(u)\mathrm{d}u = 1$. Moreover, $K(\cdot)$ is Lipschitz continuous on $[-1,1]$. As $T \to \infty$, $h \to 0$ and $Th \to \infty$.*

Assumption 3 imposes certain smoothness conditions on the functional coefficients, which are easily verifiable and can be regarded as a special case of Assumption 1. Assumption 4 is standard in the literature of nonparametric kernel estimation (cf., Li and Racine, 2007).

With these conditions in hand, we estimate $\boldsymbol{\mu}(\tau)$ by

$$\widehat{\boldsymbol{\mu}}(\tau) = \left[\sum_{t=1}^{T} K_h(\tau_t - \tau)\right]^{-1} \sum_{t=1}^{T} \boldsymbol{x}_t K_h(\tau_t - \tau), \tag{3.2}$$

and establish an asymptotic distribution in Theorem 3.1.

Theorem 3.1. *Let Assumptions 2–4 hold. If, in addition, $Th^5 \to \alpha \in [0,\infty)$, then, for $\forall \tau \in (0,1)$, as $T \to \infty$,*

$$\sqrt{Th}(\widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{\mu}(\tau)) \to_D N\left(\boldsymbol{\mu}_b(\tau), \tilde{v}_0 \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau)\right),$$

where $\boldsymbol{\mu}_b(\tau) = \frac{1}{2}\sqrt{\alpha}\tilde{c}_2 \boldsymbol{\mu}^{(2)}(\tau)$, $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau) = \sum_{j=0}^{\infty} \boldsymbol{B}_j(\tau) \sum_{j=0}^{\infty} \boldsymbol{B}_j^{\top}(\tau)$, and $\tilde{c}_2$ and $\tilde{v}_0$ have been defined in Section 1.

If $\alpha = 0$, there is no bias term involved in the asymptotic distribution of Theorem 3.1, which then falls in the usual undersmoothing scenario (Li and Racine, 2007). To establish valid inferences, both $\boldsymbol{\mu}_b(\tau)$ and $\tilde{v}_0 \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau)$ have to be estimated, and we apply both the DWB and HAC methods of

Section 2. In particular, the DWB procedure is able to handle the estimation of both $\boldsymbol{\mu}_b(\tau)$ and $\tilde{v}_0\boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau)$ simultaneously.

The DWB Method — The implementation is as follows:

1. For $\forall\tau \in (0,1)$, let $\widetilde{\boldsymbol{\mu}}(\tau)$ be defined in the same way as in (3.2) using an oversmoothing bandwidth $\widetilde{h}$, and obtain the residuals: $\widetilde{\boldsymbol{e}}_t = \boldsymbol{x}_t - \widetilde{\boldsymbol{\mu}}(\tau_t)$ for $t \geq 1$.

2. Generate $\boldsymbol{x}_t^* = \widetilde{\boldsymbol{\mu}}(\tau_t) + \boldsymbol{e}_t^*$ with $\boldsymbol{e}_t^* = \xi_t^*\widetilde{\boldsymbol{e}}_t$, in which ξ_t^*'s form an l–dependent time series satisfying $E[\xi_t^*] = 0$, $E[\xi_t^{*2}] = 1$, $E|\xi_t^*|^\delta < \infty$ for some $\delta > 2$, and $E[\xi_t^*\xi_s^*] = a\left((t-s)/l\right)$ with a kernel function $a(\cdot)$ and a tuning parameter l.

3. Use $\boldsymbol{x}_t^*$'s to construct an estimator $\widehat{\boldsymbol{\mu}}^*(\tau)$ as in (3.2).

4. Repeat Steps 2–3 J times. Let $\boldsymbol{q}_\alpha(\tau)$ be the α–quantile of the J statistics $\widehat{\boldsymbol{\mu}}^*(\tau) - \widetilde{\boldsymbol{\mu}}(\tau)$, and denote the $(1-\alpha)\cdot 100\%$ confidence interval of $\widehat{\boldsymbol{\mu}}(\tau)$ as

$$\left[\widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{q}_{1-\alpha/2}(\tau),\ \widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{q}_{\alpha/2}(\tau)\right].$$

Here, $\widetilde{h}$ is an oversmoothing bandwidth, as we shall require $h/\widetilde{h} \to 0$, where h is the same as that in (3.2). The asymptotic properties for the DWB procedure are given in Theorem 3.2 below.

Theorem 3.2. *Let $l \to \infty$, $\max\{\widetilde{h}, h/\widetilde{h}\} \to 0$ and $l \cdot \max\{1/\sqrt{Th}, \widetilde{h}^4\} \to 0$. Additionally, let $a(\cdot)$ be a symmetric kernel and Lipschitz continuous on $[-1,1]$ satisfying that $a(0) = 1$ and*

$$K_a(x) = \int_{-\infty}^{\infty} a(u)e^{-iux}\mathrm{d}u \geq 0 \quad \text{for } x \in \mathbb{R}.$$

Under the conditions of Proposition 3.1, for $\forall\tau \in (0,1)$

1. $\sup_{\boldsymbol{w}\in\mathbb{R}^d}\left|\Pr^*\left[\sqrt{Th}\left(\widehat{\boldsymbol{\mu}}^*(\tau) - \widetilde{\boldsymbol{\mu}}(\tau)\right) \leq \boldsymbol{w}\right] - \Pr\left[\sqrt{Th}\left(\widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{\mu}(\tau)\right) \leq \boldsymbol{w}\right]\right| = o_P(1)$,
2. $\liminf_{T\to\infty}\Pr\left(\boldsymbol{\mu}(\tau) \in \left[\widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{q}_{1-\alpha/2}(\tau), \widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{q}_{\alpha/2}(\tau)\right]\right) = 1-\alpha$,

where $\Pr^*$ denotes the probability measure induced by the DWB procedure.

Theorem 3.2 shows that the confidence interval of $\boldsymbol{\mu}(\tau)$ can be recovered by the empirical quantile of $\widehat{\boldsymbol{\mu}}^*(\tau) - \widetilde{\boldsymbol{\mu}}(\tau)$. Note that there is no need to deal with the bias in the DWB procedure, as the bootstrap draws generate a bias term identical to that in Theorem 3.1 (see (A.5) of Appendix A for the technical details).

The HAC Method — The HAC estimator is naturally given by:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}(\tau) = \widehat{\boldsymbol{\Xi}}_0(\tau) + \sum_{i=1}^{b}\psi\left(i/b\right)\left(\widehat{\boldsymbol{\Xi}}_i(\tau) + \widehat{\boldsymbol{\Xi}}_i^\top(\tau)\right), \tag{3.3}$$

where $\widehat{\boldsymbol{\Xi}}_i(\tau) = \left[\sum_{t=1}^{T-i} K\left(\frac{\tau_t-\tau}{h}\right)\right]^{-1} \sum_{t=1}^{T-i} \widehat{\boldsymbol{e}}_t \widehat{\boldsymbol{e}}_{t+i}^\top K\left(\frac{\tau_t-\tau}{h}\right)$ for $i \geq 0$, $\widehat{\boldsymbol{e}}_t = \boldsymbol{x}_t - \widehat{\boldsymbol{\mu}}(\tau_t)$, $\psi(\cdot)$ is a kernel function, and b is the bandwidth diverging at a relatively slow rate.

Under some mild conditions, we summarize the asymptotic property of (3.3) in the following theorem.

Theorem 3.3. *Suppose that $\psi(\cdot)$ is Lipschitz continuous, and has a compact support on $[-1, 1]$ with $\psi(0) = 1$. Additionally, let $b \to \infty$ and $b/\sqrt{Th} \to 0$. Under the conditions of Theorem 3.1, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}(\tau) = \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau) + o_P(1)$ for $\forall \tau \in (0, 1)$.*

It should be pointed out that the HAC method does not handle the bias term at all, so it only generates valid inference when $Th^5 \to 0$. To estimate the bias term in this case, one will have to employ a higher–order local polynomial approach as in Xia (1998) and Hall and Racine (2015). We no longer purse the latter in this study.

As a further application, in the next section we establish uniform consistency of nonparametric kernel estimators of the time–varying coefficients in a class of time–varying VARX models. In addition, we are able to estimate the parametric components with a $\sqrt{T}$-convergence rate for semiparametric time–varying VARX models.

4 Estimation of Time–Varying VARX Models

In this section we use the results in Section 2 to derive asymptotic properties for non– and semi–parametric estimators in a class of time–varying VARX models of the form:

$$\boldsymbol{y}_t = \sum_{j=1}^{p} \boldsymbol{A}_j(\tau_t)\boldsymbol{y}_{t-j} + \sum_{j=0}^{q} \boldsymbol{B}_j(\tau_t)\boldsymbol{x}_{t-j} + \boldsymbol{\eta}_t := \boldsymbol{Z}_t^\top \boldsymbol{\beta}(\tau_t) + \boldsymbol{\eta}_t \tag{4.1}$$

where $\boldsymbol{Z}_t = \boldsymbol{z}_t \otimes \boldsymbol{I}_d$, $\boldsymbol{z}_t = (\boldsymbol{y}_{t-1}^\top, \ldots, \boldsymbol{y}_{t-p}^\top, \boldsymbol{x}_t^\top, \boldsymbol{x}_{t-1}^\top, \ldots, \boldsymbol{x}_{t-q}^\top)^\top$, and $\boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau_t)\boldsymbol{\epsilon}_t$. Here, $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{d,t})^\top$ is a d-dimensional vector of endogenous variables, $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{m,t})^\top$ is an m-dimensional vector of exogenous variables, and both d and m are finite integers. Accordingly, $\{\boldsymbol{A}_j(\tau)\}$ and $\{\boldsymbol{B}_j(\tau)\}$ are the $d \times d$ and $d \times m$ coefficient matrices. Also, $\boldsymbol{\omega}(\tau)$ is an unknown deterministic function which has full row rank uniformly in $\tau \in [0, 1]$, and captures the heteroscedasticity over time. Obviously, we have

$$\boldsymbol{\beta}(\tau) = \text{vec}(\boldsymbol{A}(\tau), \boldsymbol{B}(\tau)), \tag{4.2}$$

where $\boldsymbol{A}(\tau) = (\boldsymbol{A}_1(\tau), \ldots, \boldsymbol{A}_p(\tau))$ and $\boldsymbol{B}(\tau) = (\boldsymbol{B}_0(\tau), \boldsymbol{B}_1(\tau), \ldots, \boldsymbol{B}_q(\tau))$.

In what follows, we are interested to estimate $\{\boldsymbol{A}_j(\tau)\}$ and $\{\boldsymbol{B}_j(\tau)\}$, and are particularly interested to adopt the nonparametric local linear approach and the semiparametric profile likelihood estimation of Fan and Huang (2005). In the following two subsections, we consider both non– and semi–parametric versions of time–varying VARX models.

4.1 Nonparametric Estimation

We start with $\boldsymbol{\beta}(\cdot)$, and assume each component of $\boldsymbol{\beta}(\cdot)$ has continuous derivatives up to the second order. When τ_t is close to τ, we then have the following approximation:

$$\boldsymbol{y}_t \simeq \boldsymbol{Z}_t^\top \boldsymbol{\beta}(\tau) + \boldsymbol{Z}_t^\top \boldsymbol{\beta}^{(1)}(\tau)(\tau_t - \tau) + \boldsymbol{\eta}_t. \tag{4.3}$$

Usually, $\{\boldsymbol{\beta}(\tau), \boldsymbol{\beta}^{(1)}(\tau)\}$ of (4.3) can be estimated by the kernel weighted least-squares criterion:

$$(\widehat{\boldsymbol{\beta}}(\tau), \widehat{\boldsymbol{\beta}}^{(1)}(\tau)) = \underset{\boldsymbol{\beta}, \boldsymbol{\beta}^{(1)}}{\operatorname{argmin}} \sum_{t=1}^{T} \left\| \boldsymbol{y}_t - \boldsymbol{Z}_t^\top (\boldsymbol{\beta} + (\tau_t - \tau)\boldsymbol{\beta}^{(1)}) \right\|^2 K_h(\tau_t - \tau). \tag{4.4}$$

Moreover, $\widehat{\boldsymbol{\beta}}(\tau)$ admits a closed-form expression as follows:

$$\widehat{\boldsymbol{\beta}}(\tau) = (\boldsymbol{I}_l, \boldsymbol{0}_l)(\boldsymbol{Z}_\tau^\top \boldsymbol{K}_\tau \boldsymbol{Z}_\tau)^{-1} \boldsymbol{Z}_\tau^\top \boldsymbol{K}_\tau \boldsymbol{y}, \tag{4.5}$$

where $l = d^2p + (q+1)md$, $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_T^\top)^\top$,

$$\boldsymbol{K}_\tau = \operatorname{diag}\{K_h(\tau_1 - \tau), \ldots, K_h(\tau_T - \tau)\} \otimes \boldsymbol{I}_d, \quad \text{and} \quad \boldsymbol{Z}_\tau = \begin{pmatrix} \boldsymbol{Z}_1^\top & \boldsymbol{Z}_1^\top \frac{\tau_1 - \tau}{h} \\ \vdots & \vdots \\ \boldsymbol{Z}_T^\top & \boldsymbol{Z}_T^\top \frac{\tau_T - \tau}{h} \end{pmatrix}.$$

Having presented the above estimators, we introduce the following assumptions for the theoretical development.

Assumption 5.

1. *The roots of* $\boldsymbol{I}_d - \boldsymbol{A}_1(\tau)L - \cdots - \boldsymbol{A}_p(\tau)L^p = \boldsymbol{0}_d$ *all lie outside the unit circle uniformly in* $\tau \in [0,1]$.

2. *Each element of* $\boldsymbol{\beta}(\tau)$ *is second–order continuously differentiable on* $[0,1]$ *and* $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta}(0)$ *for* $\tau < 0$.

3. *Suppose that*

$$\boldsymbol{x}_t = \boldsymbol{g}(\tau_t) + \sum_{j=0}^{\infty} \boldsymbol{D}_j(\tau_t)\boldsymbol{v}_{t-j} \quad \text{for} \quad t \geq 1,$$
$$\boldsymbol{x}_t = \boldsymbol{g}(0) + \sum_{j=0}^{\infty} \boldsymbol{D}_j(0)\boldsymbol{v}_{t-j} \quad \text{for} \quad t \leq 0,$$

where $\boldsymbol{g}(\cdot)$ *and* $\boldsymbol{D}_j(\cdot)$ *are* $m \times 1$ *and* $m \times m$ *respectively. Each component of* $\boldsymbol{g}(\cdot)$ *and* $\boldsymbol{D}_j(\cdot)$ *is second-order continuously differentiable on* $[0,1]$. *For* $\ell = 0, 1$, $\sup_{\tau \in [0,1]} \sum_{j=1}^{\infty} j \|\boldsymbol{D}_j^{(\ell)}(\tau)\| < \infty$.

4. *Each component of* $\boldsymbol{\omega}(\tau)$ *is second-order continuously differentiable on* $[0,1]$. *Moreover,* $\boldsymbol{\Omega}(\tau)=\boldsymbol{\omega}(\tau)\boldsymbol{\omega}(\tau)^\top$ *is positive definite uniformly in* $\tau\in[0,1]$, *and* $\boldsymbol{\omega}(\tau)=\boldsymbol{\omega}(0)$ *for* $\tau<0$.

5. *Let* $\boldsymbol{e}_t = (\boldsymbol{\epsilon}_t^\top, \boldsymbol{v}_{t+1}^\top)^\top$ *and* $\{\boldsymbol{e}_t\}_{t=-\infty}^{\infty}$ *form a sequence of martingale differences such that* $E\left(\boldsymbol{e}_t|\mathcal{F}_{t-1}\right)=0$, *where* $\mathcal{F}_t=\sigma\{\boldsymbol{e}_t,\boldsymbol{e}_{t-1},\ldots\}$. *Also, suppose that* $E(\boldsymbol{e}_t\boldsymbol{e}_t^\top|\mathcal{F}_{t-1}) = \begin{pmatrix}\boldsymbol{I}_d & \boldsymbol{\rho}_{\boldsymbol{\epsilon v}}\\ \boldsymbol{\rho}_{\boldsymbol{\epsilon v}}^\top & \boldsymbol{I}_m\end{pmatrix}$ *almost surely (a.s.), and* $\max_{t\geq 1} E\left\|\boldsymbol{e}_t\right\|^\delta<\infty$ *for some* $\delta\geq 4$.

Assumptions 5.2 is pretty standard in the literature (Li and Racine, 2007), so the discussions are omitted. Assumption 5.5 is also standard by assuming that the innovation errors follow a martingale difference structure, which is identical to those used in Phillips and Lee (2013) for example.

We now comment on the rest conditions of Assumption 5. Assumption 5.1 ensures that $\boldsymbol{y}_t$ in model (4.1) is neither a unit-root process nor an explosive process, and can be regarded as an extension of those used for the classical multivariate dynamic models (e.g., Hamilton, 1994, p. 259). Assumption 5.3 formulates a time-varying VMA(∞) process which nests many different processes as special cases as shown in Examples 1–3. Assumption 5.4 imposes the heteroscedasticity on the structure using an unknown function, and the assumptions are in the same spirit of those for $\boldsymbol{\beta}(\cdot)$.

The following theorem establishes the asymptotic properties associated with the estimation procedure of (4.5).

Theorem 4.1. *Let Assumptions 4 and 5 hold. If* $T\to\infty$, *then*

1. $\forall\tau\in(0,1)$ *we have*

$$\sqrt{Th}(\widehat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}(\tau)-\frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\beta}^{(2)}(\tau)+o_P(h^2))\to_D N\left(\mathbf{0},\widetilde{v}_0\boldsymbol{V}(\tau)\right)$$

where $\boldsymbol{V}(\tau)=\boldsymbol{\Sigma}_{\boldsymbol{z}}^{-1}(\tau)\otimes\boldsymbol{\Omega}(\tau)$ *and* $\boldsymbol{\Sigma}_{\boldsymbol{z}}(\tau)=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T E(\boldsymbol{z}_t\boldsymbol{z}_t^\top)K_h(\tau_t-\tau)$;

If, in addition, $\max_{t\geq1}E\left[\|\boldsymbol{e}_t\|^4|\mathcal{F}_{t-1}\right]<\infty$ *a.s. and* $\frac{T^{1-4/\delta}h}{\log T}\to\infty$, *then we have*

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

3. $\widehat{\boldsymbol{V}}(\tau)\to_P\boldsymbol{V}(\tau)$ *for* $\forall\tau\in[0,1]$, *where* $\widehat{\boldsymbol{V}}(\tau):=\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{z}}^{-1}(\tau)\otimes\widehat{\boldsymbol{\Omega}}(\tau)$, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{z}}(\tau)=(\frac{1}{T}\sum_{t=1}^T K_h(\tau_t-\tau))^{-1}\frac{1}{T}\sum_{t=1}^T\boldsymbol{z}_t\boldsymbol{z}_t^\top K_h(\tau_t-\tau)$, $\widehat{\boldsymbol{\Omega}}(\tau)=(\frac{1}{T}\sum_{t=1}^T K_h(\tau_t-\tau))^{-1}\frac{1}{T}\sum_{t=1}^T\widehat{\boldsymbol{\eta}}_t\widehat{\boldsymbol{\eta}}_t^\top K_h(\tau_t-\tau)$ *and* $\widehat{\boldsymbol{\eta}}_t=\boldsymbol{y}_t-\boldsymbol{Z}_t^\top\widehat{\boldsymbol{\beta}}(\tau_t)$.

4.2 Semiparametric Estimation

In this subsection, we consider a semiparametric version of model (4.1), assuming that some of $\boldsymbol{\beta}(\cdot)$ are not time–varying. Let $\boldsymbol{C}$ be a $s\times l$ selection matrix with $1\leq s<l$ such that $\boldsymbol{C}\boldsymbol{\beta}(\cdot)=\boldsymbol{c}$. In addition, let $\widetilde{\boldsymbol{C}}$ be a selection matrix collecting the elements of $\boldsymbol{\beta}(\tau)$ left out by $\boldsymbol{C}$. Thus, (4.1) can be rewritten as

$$\boldsymbol{y}_t = \boldsymbol{X}_{\boldsymbol{C},t}^\top \boldsymbol{c} + \boldsymbol{X}_{\widetilde{\boldsymbol{C}},t}^\top \boldsymbol{\theta}(\tau) + \boldsymbol{\eta}_t, \tag{4.6}$$

where $\boldsymbol{X}_{\boldsymbol{C},t} = \boldsymbol{C}\boldsymbol{Z}_t$, $\boldsymbol{X}_{\widetilde{\boldsymbol{C}},t} = \widetilde{\boldsymbol{C}}\boldsymbol{Z}_t$, and $\boldsymbol{\theta}(\tau) = \widetilde{\boldsymbol{C}}\boldsymbol{\beta}(\tau)$. The right hand side of (4.6) reduces to a semiparametric time-varying model. Using the profile likelihood estimation therein, $\boldsymbol{c}$ can be estimated by

$$\widehat{\boldsymbol{c}} = \left(\boldsymbol{X}_{\boldsymbol{C}}^\top(\boldsymbol{I}_{dT} - \boldsymbol{S})^\top(\boldsymbol{I}_{dT} - \boldsymbol{S})\boldsymbol{X}_{\boldsymbol{C}}\right)^{-1} \boldsymbol{X}_{\boldsymbol{C}}^\top(\boldsymbol{I}_{dT} - \boldsymbol{S})^\top(\boldsymbol{I}_{dT} - \boldsymbol{S})\boldsymbol{y}, \tag{4.7}$$

where $\boldsymbol{S} = (\boldsymbol{s}(\tau_1)^\top \boldsymbol{X}_{\widetilde{\boldsymbol{C}},1}, \ldots, \boldsymbol{s}(\tau_T)^\top \boldsymbol{X}_{\widetilde{\boldsymbol{C}},T})^\top$, $\boldsymbol{s}(\tau) = (\boldsymbol{I}_{l-s}, \boldsymbol{0}_{l-s})(\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^\top \boldsymbol{K}_\tau \boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau})^{-1}\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^\top \boldsymbol{K}_\tau$, and

$$\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau} = \begin{pmatrix} \boldsymbol{X}_{\widetilde{\boldsymbol{C}},1}^\top & \boldsymbol{X}_{\widetilde{\boldsymbol{C}},1}^\top \frac{\tau_1 - \tau}{h} \\ \vdots & \vdots \\ \boldsymbol{X}_{\widetilde{\boldsymbol{C}},T}^\top & \boldsymbol{X}_{\widetilde{\boldsymbol{C}},T}^\top \frac{\tau_T - \tau}{h} \end{pmatrix}.$$

Finally, $\boldsymbol{\theta}(\tau)$ of (4.6) can be estimated by

$$\widehat{\boldsymbol{\theta}}(\tau) = (\boldsymbol{I}_{l-s}, \boldsymbol{0}_{l-s})(\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^\top \boldsymbol{K}_\tau \boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau})^{-1}\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^\top \boldsymbol{K}_\tau \left(\boldsymbol{y} - \boldsymbol{X}_{\boldsymbol{C}}\widehat{\boldsymbol{c}}\right), \tag{4.8}$$

where $\boldsymbol{X}_{\boldsymbol{C}} = (\boldsymbol{X}_{\boldsymbol{C},1}, \ldots, \boldsymbol{X}_{\boldsymbol{C},T})^\top$.

Having proposed the above estimators, we introduce following assumption for the establishment of a semiparametric estimation theory.

Assumption 6. *Let* $\max_{t\geq 1} E[\|\boldsymbol{e}_t\|^4 \,|\mathcal{F}_{t-1}] < \infty$ *a.s.,* $Th^8 \to 0$, $\frac{Th^2}{(\log T)^2} \to \infty$, $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$, *and* $\delta > 4$, *where* δ *is the same as that of Assumption 5.5.*

Assumption 6 imposes more restrictive conditions on the bandwidth, and the conditional moments of the error terms. These assumptions are commonly used in the literature of semiparametric kernel estimation (e.g., Fan and Huang, 2005).

With Assumptions 5 and 6 in hand, the next theorem establishes the asymptotic distributions associated with the estimation procedure of (4.7) and (4.8).

Theorem 4.2. *Let Assumptions 4 and 6 hold, and* $T \to \infty$.

1. *For* (4.7),
$$\sqrt{T}\,(\widehat{\boldsymbol{c}} - \boldsymbol{c}) \to_D N\left(\boldsymbol{0}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}\boldsymbol{\Sigma}^{-1}\right),$$
where $\boldsymbol{\Sigma} = \int_0^1 (\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C}}}(\tau) - \boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}(\tau)\boldsymbol{\Sigma}_{\boldsymbol{X}_{\widetilde{\boldsymbol{C}}}}^{-1}(\tau)\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}^\top(\tau))\mathrm{d}\tau$, $\boldsymbol{\Delta} = \int_0^1 \boldsymbol{P}_{\boldsymbol{C}}(\tau)\left(\boldsymbol{\Sigma}_{\boldsymbol{z}}(\tau) \otimes \boldsymbol{\Omega}(\tau)\right)\boldsymbol{P}_{\boldsymbol{C}}^\top(\tau)\mathrm{d}\tau$, $\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C}}}(\tau) = \boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{Z}}(\tau)\boldsymbol{C}^\top$, $\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}(\tau) = \boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{Z}}(\tau)\widetilde{\boldsymbol{C}}^\top$, $\boldsymbol{\Sigma}_{\boldsymbol{X}_{\widetilde{\boldsymbol{C}}}}(\tau) = \widetilde{\boldsymbol{C}}\boldsymbol{\Sigma}_{\boldsymbol{Z}}(\tau)\widetilde{\boldsymbol{C}}^\top$ *and* $\boldsymbol{P}_{\boldsymbol{C}}(\tau) = \boldsymbol{C} - \boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}(\tau)\boldsymbol{\Sigma}_{\boldsymbol{X}_{\widetilde{\boldsymbol{C}}}}^{-1}(\tau)\widetilde{\boldsymbol{C}}$;

2. *For* (4.8), $\forall \tau \in (0,1)$
$$\sqrt{Th}(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau) - \frac{1}{2}h^2\tilde{c}_2\boldsymbol{\theta}^{(2)}(\tau) + o_P(h^2)) \to_D N(\boldsymbol{0}, \tilde{v}_0\boldsymbol{\Delta}_{\boldsymbol{\theta}}(\tau)),$$

where $\boldsymbol{\Delta}_{\boldsymbol{\theta}}(\tau) = \boldsymbol{\Sigma}^{-1}_{\boldsymbol{X}_{\widetilde{\boldsymbol{C}}}}(\tau)\widetilde{\boldsymbol{C}}\left(\boldsymbol{\Sigma}_{\boldsymbol{z}}(\tau)\otimes\boldsymbol{\Omega}(\tau)\right)\widetilde{\boldsymbol{C}}^\top\boldsymbol{\Sigma}^{-1}_{\boldsymbol{X}_{\widetilde{\boldsymbol{C}}}}(\tau)$.

Similar to Theorem 4.1 (3), both $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Delta}_{\boldsymbol{\theta}}(\tau)$ can be easily estimated by replacing the unknown quantities with their estimators. In the following section, we conduct numerical studies to evaluate the finite–sample performance of the proposed estimation and inferential methods.

5 Numerical Studies

In this section, we first present the details of the numerical implement in Section 5.1, and then conduct extensive simulations in Section 5.2. Since the simulation results show that the DWB approach works better numerically than the HAC method, we therefore only apply the DWB approach for our empirical analysis in Section 5.3.

5.1 Numerical Implementation

We provide some details for practical implementation when applying the results in Section 3. The Epanechnikov kernel $K(u) = 0.75(1-u^2)I(|u|\leq 1)$ is adopted throughout the numerical studies. For bandwidth selection, since the error innovations, $\boldsymbol{e}_t = \sum_{j=0}^{\infty}\boldsymbol{B}_j(\tau_t)\boldsymbol{\epsilon}_{t-j}$, involved in the time–varying VMA(∞) model are serially correlated, we use the modified cross–validation criterion proposed by Chu and Marron (1991). Specifically, it is a "leave–($2k$+1)–out" version of cross–validation, and $\widehat{h}_{mcv}$ is selected by the following minimization procedure.

$$\widehat{h}_{mcv} = \arg\min_h \sum_{t=1}^{T}\left(\boldsymbol{x}_t - \widehat{\boldsymbol{\mu}}_{k,h}(\tau_t)\right)^\top\left(\boldsymbol{x}_t - \widehat{\boldsymbol{\mu}}_{k,h}(\tau_t)\right), \tag{5.1}$$

where $\widehat{\boldsymbol{\mu}}_{k,h}(\tau) = \left[\sum_{t:|t-\tau T|>k} K\left(\frac{\tau_t-\tau}{h}\right)\right]^{-1}\sum_{t:|t-\tau T|>k}\boldsymbol{x}_t K\left(\frac{\tau_t-\tau}{h}\right)$ and $k=5$.

We then comment on the HAC and DWB methods, of which both require a bandwidth and a kernel function. For the HAC procedure, we use the Bartlett kernel $\psi(x) = (1-|x|)I(|x|\leq 1)$ (e.g., Newey and West, 1987) and the rule of thumb bandwidth $b = 0.75\cdot(T\widehat{h}_{mcv})^{1/3}$. For the DWB method, we follow the suggestions of Bühlmann (1998), Shao (2010) and Palm et al. (2011) by choosing $\widetilde{h} = c_0\cdot\widehat{h}_{mcv}^{5/9}$ with $c_0 = 2$, $a(x) = \frac{\int_{-1}^{1} w(u)w(u+|x|)\mathrm{d}u}{\int_{-1}^{1} w^2(u)\mathrm{d}u}$ with $w(u) = \frac{u}{0.43}I\left(u\in[0,0.43)\right) + I\left(u\in[0.43,0.57]\right) + \frac{1-u}{0.43}I\left(u\in(0.57,1]\right)$, and $l = 1.75\cdot(T\widehat{h}_{mcv})^{1/3}$ respectively.

5.2 Simulation Results

We first evaluate the finite sample performance of the DWB and HAC procedures presented in Section 3. Consider a multivariate time series with the following data generating process (DGP)

$$\boldsymbol{x}_t = \boldsymbol{\mu}(\tau_t) + \boldsymbol{e}_t, \quad \boldsymbol{e}_t = \boldsymbol{A}(\tau_t)\boldsymbol{e}_{t-1} + \boldsymbol{\epsilon}_t, \quad t = 1,2,...,T, \tag{5.2}$$

where $\boldsymbol{\epsilon}_t$'s are i.i.d. draws from $N(\mathbf{0}_{2\times 1}, \boldsymbol{I}_2)$, $\boldsymbol{\mu}(\tau) = [\sin(\pi\tau), \cos(\pi\tau)]^\top$, and

$$\boldsymbol{A}(\tau) = \begin{bmatrix} 0.3\exp(-0.5+\tau) & (\tau-0.5)^3 \\ (\tau-0.5)^3 & 0.3+0.3\sin(\pi\tau) \end{bmatrix}.$$

In addition, let sample size be $T \in \{200, 400, 800\}$ and conduct 1000 replications for each choice of T.

In order to evaluate the finite sample performance, we calculate the point–wise coverage rate associated with $\boldsymbol{\mu}(\cdot)$ based on the HAC estimation method and the DWB procedure with $J = 1000$ bootstrap replications, respectively. Specifically, we consider the coverage at $\tau = 0.1, \ldots, 0.9$, and use the nominal coverage 95%. For each given τ, the coverage probability is first calculated for each component of $\boldsymbol{\mu}(\cdot)$ over 1000 replications, and then we take average across the elements of $\boldsymbol{\mu}(\cdot)$. These probabilities are reported in Table 1. It can be seen that the DWB method yields better coverage probabilities, which approach 95% faster than those from the HAC method. For this reason, we will use the DWB method in the empirical study below. In addition, we conjecture that the performance of HAC method can be improved by using a bias corrected trending estimator as explained above. Since our simulation results show that the DWB approach works better than the HAC method numerically, we therefore apply the DWB approach in our empirical analysis.

Table 1: Point-wise coverage probabilities for $\boldsymbol{\mu}(\cdot)$

	Mean		Median	
T	DWB	HAC	DWB	HAC
200	0.901	0.818	0.895	0.815
400	0.901	0.833	0.896	0.829
800	0.930	0.868	0.925	0.864

We next evaluate the performance of the semiparametric profile likelihood method for the following DGP:

$$\boldsymbol{y}_t = \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \boldsymbol{B}_1(\tau)x_{t-1} + \boldsymbol{\eta}_t, \tag{5.3}$$

where $\boldsymbol{\eta}_t$'s are i.i.d. draws from $N(\mathbf{0}_{2\times 1}, \boldsymbol{I}_2)$,

$$\boldsymbol{A}_1 = \begin{bmatrix} 0.4 & -0.1 \\ -0.1 & 0.4 \end{bmatrix}$$

$$\boldsymbol{B}_1(\tau) = [2\exp(\tau-1)-1, 2\exp(\tau-1)-1]^\top,$$

in which x_t is an AR(1) process such that $x_t = 0.4x_{t-1} + v_t$ with $v_t \sim N(0,1)$. We set T to be $200, 400, 800$ and conduct 1000 replications for each choice of T. For bandwidth selection, here we use the rule of thumb bandwidth $\widehat{h} = 2.34\sqrt{1/12}T^{-1/5}$ for simplicity.

We evaluate the estimates of $\boldsymbol{A}_1$ and $\boldsymbol{B}_1(\tau)$. For each parameter of interest, we report the finite sample coverage probabilities of the confidence intervals. The nominal level is 95%. Specifically, for $\boldsymbol{B}_1(\tau)$, the coverage probability is first calculated for each functional component over the grid points $\{\tau_t : t = 1, 2, ..., T\}$, and then we further take an average across the elements of $\boldsymbol{B}_1(\tau)$. After 1000 replications, we present the averaged value of these coverage probabilities in Table 2.

As shown in Table 2, the finite sample coverage probability of $\boldsymbol{B}_1(\cdot)$ is smaller than their nominal level when $T = 200$, but are fairly close to 95% as $T = 800$. In addition, the empirical coverage probability of $\boldsymbol{A}_1$ is very close to the nominal level even when the sample size is relatively small. This result is expected since the rate of convergence on the time-invariant components can be improved to reach a parametric rate.

Table 2: Empirical coverage probabilities for $\boldsymbol{A}_1$ and $\boldsymbol{B}_1(\cdot)$

T	$\boldsymbol{A}_1$	$\boldsymbol{B}_1(\cdot)$
200	0.952	0.908
400	0.954	0.924
800	0.950	0.939

5.3 A Real Data Example

In this subsection, we infer the long–run level of inflation (i.e., trend inflation) and the natural rate of unemployment (NAIRU, which measures the frictional and structural unemployment) based on model (3.1). The trend inflation and NAIRU are of central position in setting monetary policy since the Federal Reserve Bank aims to mitigate deviations of inflation and unemployment from their long–run targets (Primiceri, 2006; Stock and Watson, 2016). The estimation is conducted in exactly the same way as in Section 5.1, so we will no longer repeat the details unless necessary.

Specifically, we estimate the time–varying VMA(∞) model (3.1) using three commonly adopted macroeconomic variables of the literature (Primiceri, 2005; Cogley et al., 2010), which are the inflation rate (measured by the 100 times the year–over–year log change in the GDP deflator), the unemployment rate and the interest rate (measured by the average value for the Federal funds rates over the quarter). Although we are not interested in the trend of interest rates, we include this variable within the system in order to capture more dynamics and be consistent with the literature. The data are quarterly observations measured at an annual rate from 1954:Q3 to 2020:Q1, which are collected from the Federal Reserve Bank of St. Louis economic database. Figure 1 plots the three variables.

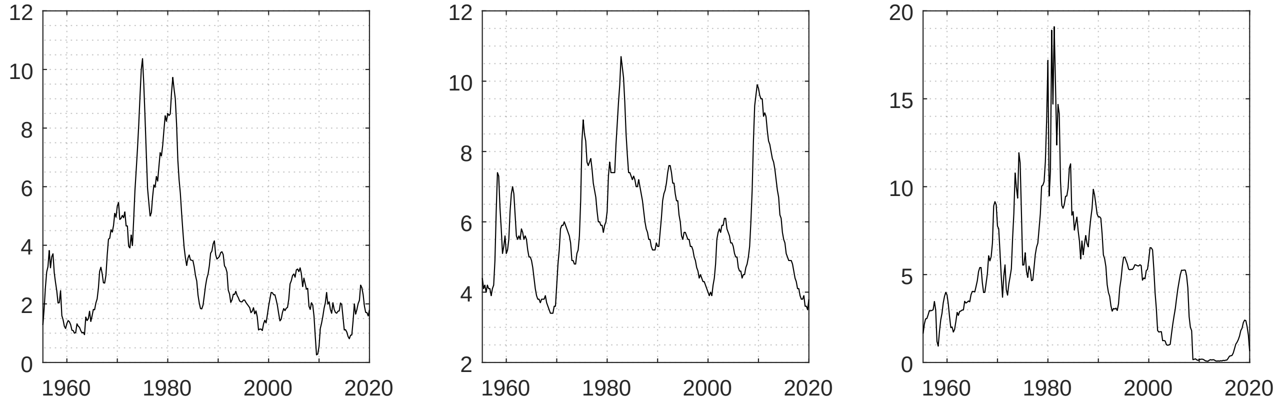

Figure 1: Plots of the inflation (left), the unemployment rate (middle) and the interest rate (right)

We investigate the trend inflation and the NAIRU. Petrova (2019) considers a Bayesian time–varying VAR(2) model, and induces the long–run mean of $\boldsymbol{x}_t$ by

$$\boldsymbol{\mu}_t = \lim_{p\to\infty} E_t(\boldsymbol{x}_{t+p}) = (\boldsymbol{I}_2 - \boldsymbol{A}_{1t} - \boldsymbol{A}_{2t})^{-1}\boldsymbol{a}_t, \tag{5.4}$$

where $\boldsymbol{a}_t$ is the intercept term, and $\boldsymbol{A}_{1t}$ and $\boldsymbol{A}_{2t}$ are the coefficient matrices. The main difference between our method and the Petrova's method is that we explicitly estimate the underlying trends of inflation and unemployment using the model (3.1).

Figure 2 plots the estimates of the trend inflation and the NAIRU (i.e., $\widehat{\boldsymbol{\mu}}(\tau)$), as well as the 95% bootstrap confidence intervals. It is obvious that the underlying trend of inflation is high in the 1970s, but decreases in the subsequent period. After the Great Moderation, the long–run level of inflation is below, but quite close to the Federal Reserve's target of 2%, which indicates that the inflation is more anchored now than in the 1970s. However, the NAIRU is less persistent and fluctuates over time. In particular, the NAIRU increases rapidly during "Second Oil Crisis" and "Global Financial Crisis".

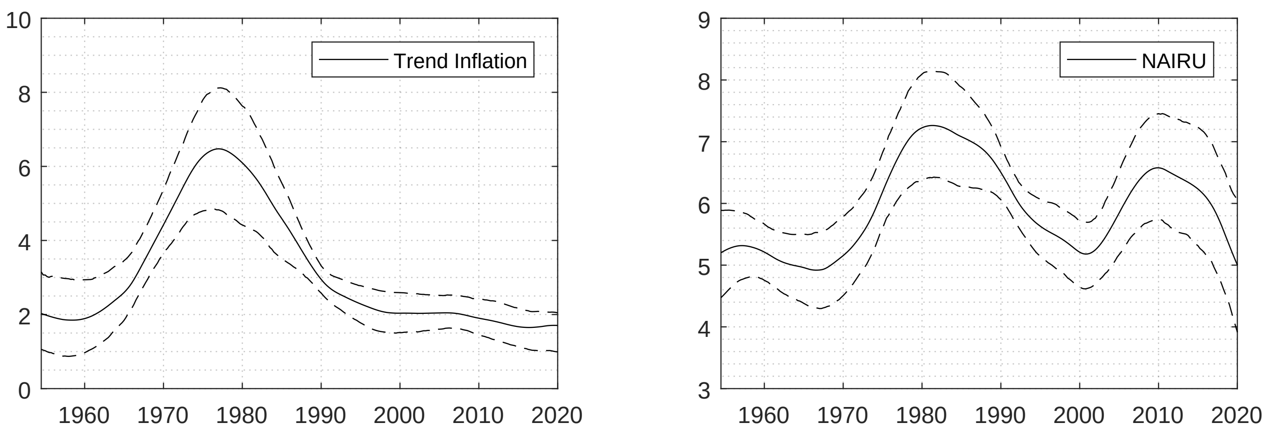

Figure 2: The estimated trends (i.e., $\boldsymbol{\mu}(\cdot)$) of inflation and unemployment as well as the associated 95% bootstrap confidence intervals

6 Conclusion

In this paper, we introduce a class of time–varying VMA(∞) processes, and derive a set of asymptotic properties accordingly. Our investigation starts with decomposing the weighted sum of time–varying VMA(∞) processes into the long–run and transitory elements, that is known as the BN decomposition (Beveridge and Nelson, 1981; Phillips and Solo, 1992). As the long–run component of the decomposition yields a martingale approximation, it ensures the feasibility of achieving a variety of asymptotics for the multivariate case, e.g., the law of large numbers, the uniform convergence, the central limit theory, the bootstrap consistency, and the long–run covariance matrix estimation. Further we show that these results can be readily applied, when establishing the inferences for many dynamic time–varying models. In the empirical study, we apply the newly proposed framework to study the long–run level of inflation and the natural rate of unemployment. We find that (1) the long–run level of inflation is more anchored now and is close to the Federal Reserve's target of two percent after the beginning of the Great Moderation period and (2) the natural rate of unemployment is less persistent and increases rapidly during "Second Oil Crisis" and "Global Financial Crisis".

Acknowledgements

The authors of this paper would like to thank George Athanasopoulos, Rainer Dahlhaus, David Frazier, Oliver Linton, Gael Martin, Peter C. B. Phillips and Wei Biao Wu for their constructive comments on earlier versions of this paper. Gao acknowledges financial support from the Australian Research Council Discovery Grants Program under Grant Number: DP170104421. Peng also acknowledges the Australian Research Council Discovery Grants Program for its financial support under Grant Number DP210100476.

Appendix A

In this Appendix, we first present several preliminary lemmas in Appendix A.1, which are helpful to the development of the main results. We then prove the main results on the time–varying VMA(∞) process from Sections 2–3 in Appendix A.2. In what follows, M and $O(1)$ always stand for bounded constants, and may be different at each appearance.

A.1 Preliminary Lemmas

Lemma A.1. *Suppose $\{Z_t, \mathcal{F}_t\}$ is a martingale difference sequence, $S_T = \sum_{t=1}^T Z_t$, $U_T = \sum_{t=1}^T Z_t^2$ and $s_T^2 = E(U_T^2) = E(S_T^2)$. If $s_T^{-2} U_T^2 \to_P 1$ and $\sum_{t=1}^T E[Z_{T,t}^2 I(|Z_{T,t}| > \nu)] \to 0$ for any $\nu > 0$ with $Z_{T,t} = s_T^{-1} Z_t$, then as $T \to \infty$, $s_T^{-1} S_T \to_D N(0, 1)$.*

Lemma A.1 is Corollary 3.1 of Hall and Heyde (1980).

Lemma A.2. *Let $\{Z_t, \mathcal{F}_t\}$ be a martingale difference sequence. Suppose that $|Z_t| \leq M$ for a constant M, $t = 1, \ldots, T$. Let $V_T = \sum_{t=1}^T \mathrm{Var}\,(Z_t|\mathcal{F}_{t-1}) \leq V$ for some $V > 0$. Then for any given $\nu > 0$,*

$$\Pr\left(\left|\sum_{t=1}^T Z_t\right| > \nu\right) \leq \exp\left\{-\frac{\nu^2}{2(V + M\nu)}\right\}.$$

Lemma A.2 is Proposition 2.1 of Freedman (1975).

Lemma A.3. *The following algebraic decompositions hold true.*

1. $\mathbb{B}_t(L) = \sum_{j=0}^\infty \boldsymbol{B}_{j,t}L^j$ *can be decomposed as* $\mathbb{B}_t(L) = \mathbb{B}_t(1) - (1-L)\widetilde{\mathbb{B}}_t(L)$, *where* $\widetilde{\mathbb{B}}_t(L) = \sum_{j=0}^\infty \widetilde{\boldsymbol{B}}_{j,t}L^j$ *and* $\widetilde{\boldsymbol{B}}_{j,t} = \sum_{k=j+1}^\infty \boldsymbol{B}_{k,t}$.

2. $\mathbb{B}_t^r(L) = \sum_{j=0}^\infty (\boldsymbol{B}_{j+r,t} \otimes \boldsymbol{B}_{j,t})L^j$ *can be decomposed as* $\mathbb{B}_t^r(L) = \mathbb{B}_t^r(1) - (1-L)\widetilde{\mathbb{B}}_t^r(L)$, *where* $\widetilde{\mathbb{B}}_t^r(L) = \sum_{j=0}^\infty \widetilde{\boldsymbol{B}}_{j,t}^r L^j$ *and* $\widetilde{\boldsymbol{B}}_{j,t}^r = \sum_{k=j+1}^\infty (\boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t})$.

In addition, let Assumption 1 hold, then

3. $\max_{t\geq 1} \sum_{j=0}^\infty \|\widetilde{\boldsymbol{B}}_{j,t}\| < \infty$;

4. $\limsup_{T\to\infty} \sum_{t=1}^{T-1} \|\widetilde{\mathbb{B}}_{t+1}(1) - \widetilde{\mathbb{B}}_t(1)\| < \infty$;

5. $\max_{t\geq 1} \sum_{j=0}^\infty \|\widetilde{\boldsymbol{B}}_{j,t}^r\| < \infty$;

6. $\max_{t\geq 1} \sum_{r=1}^\infty \|\widetilde{\mathbb{B}}_t^r(1)\| < \infty$;

7. $\limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{r=0}^\infty \|\widetilde{\mathbb{B}}_{t+1}^r(1) - \widetilde{\mathbb{B}}_t^r(1)\| < \infty$.

Lemma A.4. *Let Assumptions 1 and 2 hold, and let $\{\boldsymbol{W}_{T,t}(\cdot)\}_{t=1}^T$ be a sequence of $m \times d$ matrices of functions, where $m \geq 1$ is fixed, and each functional component is Lipschitz continuous and defined on a compact set $[a, b]$. Moreover, suppose that (1) $\sup_{\tau\in[a,b]} \sum_{t=1}^T \|\boldsymbol{W}_{T,t}(\tau)\| = O(1)$, and (2) $T^{\frac{2}{\delta}} d_T \log T \to 0$, where $d_T = \sup_{\tau\in[a,b],t\geq 1} \|\boldsymbol{W}_{T,t}(\tau)\|$. As $T \to \infty$, $\sup_{\tau\in[a,b]} \left\|\sum_{t=1}^T \boldsymbol{W}_{T,t}(\tau)\mathbb{B}_t(1)\boldsymbol{\epsilon}_t\right\| = O_P\left(\sqrt{d_T \log T}\right)$.*

Lemma A.5. *Let the conditions of Lemma A.4 hold. Suppose $T^{\frac{4}{\delta}} d_T \log T \to 0$, $\max_{t\geq 1} E[\|\boldsymbol{\epsilon}_t\|^4|\mathcal{F}_{t-1}] < \infty$ a.s. and $\sup_{\tau\in[a,b]} \sum_{t=1}^{T-1} \|\boldsymbol{W}_{T,t+1}(\tau) - \boldsymbol{W}_{T,t}(\tau)\| = O(d_T)$. As $T \to \infty$*

1. $\sup_{\tau\in[a,b]} \left\|\sum_{t=1}^T (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(\tau))\, \mathbb{B}_t^0(1) \left(\mathrm{vec}[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top] - \mathrm{vec}[\boldsymbol{I}_d]\right)\right\| = O_P\left(\sqrt{d_T \log T}\right)$;

2. $\sup_{\tau\in[a,b]} \left\|\sum_{t=1}^T (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(\tau))\, \boldsymbol{\zeta}_t\boldsymbol{\epsilon}_t\right\| = O_P\left(\sqrt{d_T \log T}\right)$;

where $\boldsymbol{\zeta}_t = \sum_{r=1}^\infty \sum_{s=0}^\infty \{\boldsymbol{B}_{s+r,t}\boldsymbol{\epsilon}_{t-r}\} \otimes \boldsymbol{B}_{s,t}$.

Lemma A.6. *Under Assumption 5, there exists a time–varying VMA(∞) process*

$$\widetilde{\boldsymbol{y}}_t = \boldsymbol{\mu}(\tau_t) + \sum_{j=0}^\infty \boldsymbol{D}_j^{\boldsymbol{\epsilon}}(\tau_t)\boldsymbol{\epsilon}_{t-j} + \sum_{j=0}^\infty \boldsymbol{D}_j^{\boldsymbol{v}}(\tau_t)\boldsymbol{v}_{t-j}$$

such that $\max_{t\geq 1}\{E\,\|\boldsymbol{y}_t - \widetilde{\boldsymbol{y}}_t\|^\delta\}^{1/\delta} = O(T^{-1})$, where

$$\boldsymbol{\mu}(\tau) = \sum_{j=0}^\infty \sum_{l=0}^q \boldsymbol{\Psi}_j(\tau)\boldsymbol{B}_l(\tau)\boldsymbol{g}(\tau), \quad \boldsymbol{\Psi}_j(\tau) = \boldsymbol{J}\boldsymbol{\Phi}^j(\tau)\boldsymbol{J}^\top,$$

$$\boldsymbol{\Phi}(\tau)=\begin{pmatrix}\boldsymbol{A}_1(\tau) & \cdots & \boldsymbol{A}_{p-1}(\tau) & \boldsymbol{A}_p(\tau)\\ \boldsymbol{I}_d & \cdots & \boldsymbol{0}_d & \boldsymbol{0}_d\\ \vdots & \ddots & \vdots & \vdots\\ \boldsymbol{0}_d & \cdots & \boldsymbol{I}_d & \boldsymbol{0}_d\end{pmatrix}, \quad \boldsymbol{J}=\left[\boldsymbol{I}_d,\boldsymbol{0}_{d\times d(p-1)}\right],$$

$$\boldsymbol{D}_j^{\boldsymbol{\epsilon}}(\tau)=\boldsymbol{\Psi}_j(\tau)\boldsymbol{\omega}(\tau), \quad \boldsymbol{D}_j^{\boldsymbol{v}}(\tau)=\sum_{b=\max(0,j-q)}^{j}\boldsymbol{D}_{b,j-b}^{\boldsymbol{v}}(\tau),$$

$$\boldsymbol{D}_{j,l}^{\boldsymbol{v}}(\tau)=\sum_{k=0}^{j}\boldsymbol{\Psi}_k(\tau)\boldsymbol{B}_l(\tau)\boldsymbol{C}_{j-k}(\tau).$$

Moreover, $\widetilde{\boldsymbol{y}}_t$ and $\boldsymbol{x}_t$ admit the following expression

$$\begin{pmatrix}\widetilde{\boldsymbol{y}}_t\\ \boldsymbol{x}_t\end{pmatrix}=\begin{pmatrix}\boldsymbol{\mu}(\tau_t)\\ \boldsymbol{g}(\tau_t)\end{pmatrix}+\sum_{j=0}^{\infty}\boldsymbol{D}_j(\tau_t)\begin{pmatrix}\boldsymbol{\epsilon}_{t-j}\\ \boldsymbol{v}_{t+1-j}\end{pmatrix},$$

where $\boldsymbol{D}_j(\tau)=\begin{pmatrix}\boldsymbol{D}_j^{\boldsymbol{\epsilon}}(\tau) & \boldsymbol{D}_{j-1}^{\boldsymbol{v}}(\tau)\\ \boldsymbol{0} & \boldsymbol{C}_{j-1}(\tau)\end{pmatrix}$, *and* $\boldsymbol{D}_j^{\boldsymbol{v}}(\tau)=0$ *and* $\boldsymbol{C}_j(\tau)=0$ *for* $j<0$*. Here,* $\boldsymbol{D}_j(\cdot)$*'s satisfy the same conditions as those in Assumption 3.*

Lemma A.7. *Suppose Assumptions 4–6 hold. As* $T\to\infty$*,*

1. $\sup_{\tau\in[h,1-h]}\left\|\frac{1}{T}\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^{\top}\boldsymbol{K}_\tau\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}-\boldsymbol{\Sigma}_{\boldsymbol{X}_{\widetilde{\boldsymbol{C}}}}(\tau)\otimes\boldsymbol{\Lambda}_1\right\|=O_P\left(h^2+\sqrt{\frac{\log T}{Th}}\right)$, *where* $\boldsymbol{\Lambda}_1=\mathrm{diag}(\tilde{c}_0,\tilde{c}_2)$*;*
2. $\sup_{\tau\in[h,1-h]}\left\|\frac{1}{T}\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^{\top}\boldsymbol{K}_\tau\boldsymbol{X}_{\boldsymbol{C},\tau}-\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}^{\top}(\tau)\otimes\boldsymbol{\Lambda}_2\right\|=O_P\left(h^2+\sqrt{\frac{\log T}{Th}}\right)$, *where* $\boldsymbol{\Lambda}_2=[\tilde{c}_0,0]^{\top}$*;*
3. $\sup_{\tau\in[0,1]}\left\|\frac{1}{T}\boldsymbol{Z}_\tau\boldsymbol{K}_\tau\boldsymbol{\eta}\right\|=O_P\left(\sqrt{\frac{\log T}{Th}}\right)$, *where* $\boldsymbol{\eta}=(\boldsymbol{\eta}_1,\ldots,\boldsymbol{\eta}_T)^{\top}$.

Lemma A.8. *Suppose Assumptions 4–6 hold. As* $T\to\infty$*,*

1. $\frac{1}{T}\boldsymbol{X}_{\boldsymbol{C}}^{\top}(\boldsymbol{I}_{dT}-\boldsymbol{S})^{\top}(\boldsymbol{I}_{dT}-\boldsymbol{S})\boldsymbol{X}_{\boldsymbol{C}}=\boldsymbol{\Sigma}+o_P(1)$,
   where $\boldsymbol{\Sigma}=\int_0^1\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C}}}(\tau)\mathrm{d}\tau-\int_0^1\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}(\tau)\boldsymbol{\Sigma}_{\boldsymbol{X}_{\widetilde{\boldsymbol{C}}}}^{-1}(\tau)\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}^{\top}(\tau)\mathrm{d}\tau$*;*
2. $\boldsymbol{X}_{\boldsymbol{C}}^{\top}(\boldsymbol{I}_{dT}-\boldsymbol{S})^{\top}(\boldsymbol{I}_{dT}-\boldsymbol{S})\widetilde{\boldsymbol{X}}=o_P(\sqrt{T})$, *where* $\widetilde{\boldsymbol{X}}=([\boldsymbol{X}_{\widetilde{\boldsymbol{C}},1}^{\top}\boldsymbol{\theta}(\tau_1)]^{\top},\ldots,[\boldsymbol{X}_{\widetilde{\boldsymbol{C}},T}^{\top}\boldsymbol{\theta}(\tau_T)]^{\top})^{\top}$.

A.2 Proofs of the Main Results

Proof of Lemma 2.1.

By the BN decomposition in Lemma A.3, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}(\boldsymbol{x}_t-E(\boldsymbol{x}_t))=\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}\mathbb{B}_t(1)\boldsymbol{\epsilon}_t+\frac{1}{\sqrt{T}}\widetilde{\mathbb{B}}_1(L)\boldsymbol{\epsilon}_0-\frac{1}{\sqrt{T}}\widetilde{\mathbb{B}}_{\lfloor Tr\rfloor}(L)\boldsymbol{\epsilon}_{\lfloor Tr\rfloor}$$
$$+\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor-1}\left(\widetilde{\mathbb{B}}_{t+1}(L)-\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\epsilon}_t:=\boldsymbol{I}_{T,1}+\boldsymbol{I}_{T,2}+\boldsymbol{I}_{T,3}+\boldsymbol{I}_{T,4}.$$

By the usual functional central limit theory for martingale difference sequences, we have

$$\boldsymbol{\Sigma}^{-1/2}(r)\boldsymbol{I}_{T,1} \to_D \boldsymbol{W}(r).$$

In addition, we have $\boldsymbol{I}_{T,2} \to_P 0$ uniformly over $r \in [0,1]$ as $E\left\|\widetilde{\mathbb{B}}_1(L)\boldsymbol{\epsilon}_0\right\| < \infty$ by Lemma A.3.

For $\boldsymbol{I}_{T,3}$, we need to show that

$$\sup_{r\in[0,1]}\left\|\frac{1}{\sqrt{T}}\widetilde{\mathbb{B}}_{\lfloor Tr\rfloor}(L)\boldsymbol{\epsilon}_{\lfloor Tr\rfloor}\right\| \to_P 0$$

which holds if $\max_{1\le t\le T} T^{-1}\|\widetilde{\mathbb{B}}_t(L)\boldsymbol{\epsilon}_t\|^2 \to_P 0$. This is equivalent to show for any $\nu > 0$

$$\frac{1}{T}\sum_{t=1}^{T} E\left[\|\widetilde{\mathbb{B}}_t(L)\boldsymbol{\epsilon}_t\|^2 I(\|\widetilde{\mathbb{B}}_t(L)\boldsymbol{\epsilon}_t\|^2 > T\nu)\right] \to 0.$$

Similar to the proofs of Lemma 2.4, this is satisfied due to

$$\{E\|\widetilde{\mathbb{B}}_t(L)\boldsymbol{\epsilon}_t\|^\delta\}^{1/\delta} \le M\sum_{j=1}^{\infty}\|\widetilde{\boldsymbol{B}}_{j,t}\| < \infty.$$

Finally, for $I_{T,4}$, as $E\left[\sum_{t=1}^{T-1}\left\|\left(\widetilde{\mathbb{B}}_{t+1}(L)-\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\epsilon}_t\right\|\right] < \infty$ by Lemma A.3, we have

$$\begin{aligned}\sup_{r\in[0,1]}\|I_{T,4}\| &\le \sup_{r\in[0,1]}\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor-1}\left\|\left(\widetilde{\mathbb{B}}_{t+1}(L)-\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\epsilon}_t\right\| \\ &\le \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\left\|\left(\widetilde{\mathbb{B}}_{t+1}(L)-\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\epsilon}_t\right\| = O_P(1/\sqrt{T}).\end{aligned}$$

The proof is now completed.

□

Proof of Proposition 2.1.

(1). Start from Example 1. Let ρ denote the largest eigenvalue of $\boldsymbol{\Phi}_t$ uniformly over t. Then, $\rho < 1$ by the condition in Proposition 2.1. Similar to the proof of Proposition 2.4 in Dahlhaus and Polonik (2009), we have $\max_{t\ge1}\|\prod_{i=0}^{j-1}\boldsymbol{\Phi}_{t-i}\| \le M\rho^j$, which yields that

$$\max_{t\ge1}\sum_{j=1}^{\infty} j\,\|\boldsymbol{B}_{j,t}\| = \max_{t\ge1}\sum_{j=1}^{\infty} j\left\|\boldsymbol{J}\prod_{i=0}^{j-1}\boldsymbol{\Phi}_{t-i}\boldsymbol{J}^\top\right\| \le M\sum_{j=1}^{\infty} j\rho^j = O(1).$$

In addition, for any conformable matrices $\{\boldsymbol{A}_i\}$ and $\{\boldsymbol{B}_i\}$, since

$$\prod_{i=1}^{r}\boldsymbol{A}_i - \prod_{i=1}^{r}\boldsymbol{B}_i = \sum_{j=1}^{r}\left(\prod_{k=1}^{j-1}\boldsymbol{A}_k\right)(\boldsymbol{A}_j-\boldsymbol{B}_j)\left(\prod_{k=j+1}^{r}\boldsymbol{B}_k\right),$$

we then obtain that

$$\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty} j\left\|\boldsymbol{J}\left(\prod_{i=0}^{j-1}\boldsymbol{\Phi}_{t+1-i}-\prod_{i=0}^{j-1}\boldsymbol{\Phi}_{t-i}\right)\boldsymbol{J}^\top\right\|$$

$$
\begin{aligned}
&= \limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{j=1}^{\infty} j \left\| \boldsymbol{J} \sum_{m=1}^{j} \left(\prod_{k=1}^{m-1} \boldsymbol{\Phi}_{t+2-k} \right) (\boldsymbol{\Phi}_{t+2-m} - \boldsymbol{\Phi}_{t+1-m}) \left(\prod_{k=m}^{j} \boldsymbol{\Phi}_{t+1-k} \right) \boldsymbol{J}^\top \right\| \\
&\leq M \sum_{j=1}^{\infty} j^2 \rho^{j-1} \limsup_{T\to\infty} \sum_{t=1}^{T-1} \|\boldsymbol{\Phi}_{t+1} - \boldsymbol{\Phi}_t\| = O(1)
\end{aligned}
$$

given the condition in Proposition 2.1.

Consider Example 2. Similar to Example 1,

$$
\max_{t\geq 1} \sum_{b=1}^{\infty} b \|\boldsymbol{D}_{b,t}\| \leq M \max_{t\geq 1} \sum_{b=1}^{\infty} b \sum_{j=b-q}^{b} \|\boldsymbol{B}_{j,t}\| \leq M \sum_{b=1}^{\infty} b\rho^b = O(1).
$$

In addition,

$$
\begin{aligned}
&\limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{b=1}^{\infty} b \|\boldsymbol{D}_{b,t+1} - \boldsymbol{D}_{b,t}\| \\
&\leq \limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{b=1}^{\infty} b \sum_{j=\max(0,b-q)}^{b} \|\boldsymbol{B}_{j,t+1} - \boldsymbol{B}_{j,t}\| \|\boldsymbol{\Theta}_{b-j,t+1-j}\| \\
&\quad + \limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{b=1}^{\infty} b \sum_{j=\max(0,b-q)}^{b} \|\boldsymbol{B}_{j,t}\| \|\boldsymbol{\Theta}_{b-j,t+1-j} - \boldsymbol{\Theta}_{b-j,t-j}\| \\
&\leq \max_{m,t} \|\boldsymbol{\Theta}_{m,t}\| \cdot q \cdot \limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{b=1}^{\infty} b \|\boldsymbol{B}_{b,t+1} - \boldsymbol{B}_{b,t}\| \\
&\quad + \left(\max_{t\geq 1} \sum_{b=1}^{\infty} b \sum_{j=\max(0,b-q)}^{b} \|\boldsymbol{B}_{j,t}\| \right) \cdot \left(\max_{m} \limsup_{T\to\infty} \sum_{t=1}^{T-1} \|\boldsymbol{\Theta}_{m,t+1} - \boldsymbol{\Theta}_{m,t}\| \right) = O(1).
\end{aligned}
$$

(2). By part (1) and the condition of Proposition 2.1, it suffices to show that $\max_{t\geq 1} \sum_{j=1}^{\infty} j \|\boldsymbol{D}_{j,t}\| < \infty$ and $\limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{j=1}^{\infty} j \|\boldsymbol{D}_{j,t+1} - \boldsymbol{D}_{j,t}\| < \infty$. Write

$$
\begin{aligned}
&\max_{t\geq 1} \sum_{j=1}^{\infty} j \|\boldsymbol{D}_{j,t}\| \leq M \max_{t\geq 1} \sum_{j=1}^{\infty} j \sum_{k=0}^{j} \|\boldsymbol{B}_{k,t}\| \|\boldsymbol{C}_{j-k,t-k}\| \\
&= M \max_{t\geq 1} \sum_{k=0}^{\infty} \sum_{j=k}^{\infty} j \|\boldsymbol{B}_{k,t}\| \|\boldsymbol{C}_{j-k,t-k}\| = M \max_{t\geq 1} \sum_{k=0}^{\infty} \|\boldsymbol{B}_{k,t}\| \sum_{j=0}^{\infty} (k+j) \|\boldsymbol{C}_{j,t-k}\| \\
&= M \max_{t\geq 1} \sum_{j=0}^{\infty} j \|\boldsymbol{B}_{j,t}\| \sum_{k=1}^{\infty} \|\boldsymbol{C}_{k,t-j}\| + M \max_{t\geq 1} \sum_{k=0}^{\infty} \|\boldsymbol{B}_{k,t}\| \sum_{j=0}^{\infty} j \|\boldsymbol{C}_{j,t-k}\| = O(1).
\end{aligned}
$$

In addition,

$$
\begin{aligned}
&\limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{j=1}^{\infty} j \|\boldsymbol{D}_{j,t+1} - \boldsymbol{D}_{j,t}\| \leq \limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{j=1}^{\infty} j \sum_{k=0}^{j} \|\boldsymbol{B}_{k,t}\| \cdot \|\boldsymbol{\Theta}_{t-k}\| \cdot \|\boldsymbol{C}_{j-k,t+1-k} - \boldsymbol{C}_{j-k,t-k}\| \\
&+ \limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{j=1}^{\infty} j \sum_{k=0}^{j} \|\boldsymbol{B}_{k,t}\| \cdot \|\boldsymbol{\Theta}_{t+1-k} - \boldsymbol{\Theta}_{t-k}\| \cdot \|\boldsymbol{C}_{j-k,t+1-k}\| \\
&+ \limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{j=1}^{\infty} j \sum_{k=0}^{j} \|\boldsymbol{B}_{k,t+1} - \boldsymbol{B}_{k,t}\| \cdot \|\boldsymbol{\Theta}_{t+1-k}\| \cdot \|\boldsymbol{C}_{j-k,t+1-k}\| := I_{T,1} + I_{T,2} + I_{T,3}.
\end{aligned}
$$

We only show that $I_{T,1}$ is bounded below, as the proofs of $I_{T,2}$ and $I_{T,3}$ can be established similarly.

$$
\begin{aligned}
I_{T,1} &\leq \max_{t\geq 1}\|\boldsymbol{\Theta}_t\|\cdot\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty} j\sum_{k=0}^{j}\|\boldsymbol{B}_{k,t}\|\cdot\|\boldsymbol{C}_{j-k,t+1-k}-\boldsymbol{C}_{j-k,t-k}\|\\
&= \max_{t\geq 1}\|\boldsymbol{\Theta}_t\|\cdot\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{k=0}^{\infty}\sum_{j=k}^{\infty} j\,\|\boldsymbol{B}_{k,t}\|\cdot\|\boldsymbol{C}_{j-k,t+1-k}-\boldsymbol{C}_{j-k,t-k}\|\\
&= \max_{t\geq 1}\|\boldsymbol{\Theta}_t\|\cdot\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{k=0}^{\infty}\|\boldsymbol{B}_{k,t}\|\sum_{j=0}^{\infty}(j+k)\,\|\boldsymbol{C}_{j,t+1-k}-\boldsymbol{C}_{j,t-k}\|\\
&\leq \max_{t\geq 1}\|\boldsymbol{\Theta}_t\|\cdot\max_{t\geq 1}\sum_{k=0}^{\infty}\|\boldsymbol{B}_{k,t}\|\cdot\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=0}^{\infty} j\,\|\boldsymbol{C}_{j,t+1-k}-\boldsymbol{C}_{j,t-k}\|\\
&\quad+\max_{t\geq 1}\|\boldsymbol{\Theta}_t\|\cdot\max_{t\geq 1}\sum_{k=0}^{\infty}k\,\|\boldsymbol{B}_{k,t}\|\cdot\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=0}^{\infty}\|\boldsymbol{C}_{j,t+1-k}-\boldsymbol{C}_{j,t-k}\| = O(1).
\end{aligned}
$$

The proof is now completed. □

Proof of Lemma 2.2.

By Lemma A.3, we have

$$\boldsymbol{x}_t = \boldsymbol{\mu}_t + \mathbb{B}_t(1)\boldsymbol{\epsilon}_t + \widetilde{\mathbb{B}}_t(L)\boldsymbol{\epsilon}_{t-1} - \widetilde{\mathbb{B}}_t(L)\boldsymbol{\epsilon}_t,$$

which yields

$$
\begin{aligned}
&\sum_{t=1}^{T}\boldsymbol{W}_{T,t}\left(\boldsymbol{x}_t - E(\boldsymbol{x}_t)\right) = \sum_{t=1}^{T}\boldsymbol{W}_{T,t}\mathbb{B}_t(1)\boldsymbol{\epsilon}_t + \boldsymbol{W}_{T,1}\widetilde{\mathbb{B}}_1(L)\boldsymbol{\epsilon}_0 - \boldsymbol{W}_{T,T}\widetilde{\mathbb{B}}_T(L)\boldsymbol{\epsilon}_T\\
&+\sum_{t=1}^{T-1}\left(\boldsymbol{W}_{T,t+1}\widetilde{\mathbb{B}}_{t+1}(L) - \boldsymbol{W}_{T,t}\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\epsilon}_t := \ \boldsymbol{I}_{T,1}+\boldsymbol{I}_{T,2}+\boldsymbol{I}_{T,3}+\boldsymbol{I}_{T,4}.
\end{aligned}
$$

For $\boldsymbol{I}_{T,1}$, by Assumption 2, we have

$$
\begin{aligned}
&E\left\|\sum_{t=1}^{T}\boldsymbol{W}_{T,t}\mathbb{B}_t(1)\boldsymbol{\epsilon}_t\right\|^2 = \mathrm{tr}\left(\sum_{t=1}^{T}\boldsymbol{W}_{T,t}\mathbb{B}_t(1)E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top)\mathbb{B}_t^\top(1)\boldsymbol{W}_{T,t}^\top\right)\\
&\leq M\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}\|^2 \leq M\max_{t\geq 1}\|\boldsymbol{W}_{T,t}\|\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}\| = O(d_T).
\end{aligned}
$$

Hence, $\|\boldsymbol{I}_{T,1}\| = O_P(\sqrt{d_T})$.

Also, $\|\boldsymbol{I}_{T,2}\| = O_P(d_T)$ and $\|\boldsymbol{I}_{T,3}\| = O_P(d_T)$, since $\max_{t\geq 1}\|\boldsymbol{W}_{T,t}\| = O(d_T)$, $E\|\widetilde{\mathbb{B}}_1(L)\boldsymbol{\epsilon}_0\| < \infty$ and $E\|\widetilde{\mathbb{B}}_T(L)\boldsymbol{\epsilon}_t\| < \infty$ by Lemma A.3.

For $\boldsymbol{I}_{T,4}$,

$$
\begin{aligned}
&\sum_{t=1}^{T-1}\left(\boldsymbol{W}_{T,t+1}\widetilde{\mathbb{B}}_{t+1}(L) - \boldsymbol{W}_{T,t}\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\epsilon}_t\\
&= \sum_{t=1}^{T-1}\left(\boldsymbol{W}_{T,t+1} - \boldsymbol{W}_{T,t}\right)\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\epsilon}_t + \sum_{t=1}^{T-1}\boldsymbol{W}_{T,t}\left(\widetilde{\mathbb{B}}_{t+1}(L) - \widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\epsilon}_t. \qquad \text{(A.1)}
\end{aligned}
$$

Note that for the first term on the right hand side of (A.1),

$$E\left\|\sum_{t=1}^{T-1}\left(\boldsymbol{W}_{T,t+1}-\boldsymbol{W}_{T,t}\right)\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\epsilon}_t\right\| \le \max_{t\ge1}E\left\|\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\epsilon}_t\right\|\cdot\sum_{t=1}^{T-1}\|\boldsymbol{W}_{T,t+1}-\boldsymbol{W}_{T,t}\|=O(d_T)$$

by Lemma A.3 and the conditions on $\boldsymbol{W}_{T,t}$. For the second term on the right hand side of (A.1), write

$$\begin{aligned}E\left\|\sum_{t=1}^{T-1}\boldsymbol{W}_{T,t}\left(\widetilde{\mathbb{B}}_{t+1}(L)-\widetilde{\mathbb{B}}_{t}(L)\right)\boldsymbol{\epsilon}_t\right\| &\le \max_{t\ge1}E\|\boldsymbol{\epsilon}_t\|\cdot\max_{t\ge1}\|\boldsymbol{W}_{T,t}\|\sum_{t=1}^{T-1}\|\widetilde{\mathbb{B}}_{t+1}(1)-\widetilde{\mathbb{B}}_{t}(1)\|\\ &\le M\max_{t\ge1}\|\boldsymbol{W}_{T,t}\|\sum_{t=1}^{T-1}\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}\|\boldsymbol{B}_{j,t+1}-\boldsymbol{B}_{j,t}\|\\ &= M\max_{t\ge1}\|\boldsymbol{W}_{T,t}\|\sum_{t=1}^{T-1}\sum_{j=1}^{\infty}j\,\|\boldsymbol{B}_{j,t+1}-\boldsymbol{B}_{j,t}\|=O(d_T).\end{aligned}$$

Thus, we have proved that $\|\sum_{t=1}^{T}\boldsymbol{W}_{T,t}\left(\boldsymbol{x}_t-E(\boldsymbol{x}_t)\right)\|=O_P(\sqrt{d_T})$.

We now prove $\|\sum_{t=1}^{T}\boldsymbol{W}_{T,t}\left(\boldsymbol{x}_t\boldsymbol{x}_{t+p}^\top-E\left(\boldsymbol{x}_t\boldsymbol{x}_{t+p}^\top\right)\right)\|=O_P(\sqrt{d_T})$. Start from $p=0$ and write

$$\begin{aligned}\boldsymbol{x}_t\boldsymbol{x}_t^\top &= \boldsymbol{\mu}_t\boldsymbol{\mu}_t^\top+\boldsymbol{\mu}_t\sum_{j=0}^{\infty}\boldsymbol{\epsilon}_{t-j}^\top\boldsymbol{B}_{j,t}^\top+\sum_{j=0}^{\infty}\boldsymbol{B}_{j,t}\boldsymbol{\epsilon}_{t-j}\boldsymbol{\mu}_t^\top+\sum_{j=0}^{\infty}\boldsymbol{B}_{j,t}\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-j}^\top\boldsymbol{B}_{j,t}^\top\\ &\quad+\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\boldsymbol{B}_{j,t}\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-j-r}^\top\boldsymbol{B}_{j+r,t}^\top+\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}\boldsymbol{B}_{j+r,t}\boldsymbol{\epsilon}_{t-j-r}\boldsymbol{\epsilon}_{t-j}^\top\boldsymbol{B}_{j,t}^\top,\end{aligned}$$

which yields

$$\begin{aligned}&\mathrm{vec}\left[\boldsymbol{W}_{T,t}\left(\boldsymbol{x}_t\boldsymbol{x}_t^\top-E\left(\boldsymbol{x}_t\boldsymbol{x}_t^\top\right)\right)\right]\\ &=(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t})\sum_{j=0}^{\infty}(\boldsymbol{B}_{j,t}\otimes\boldsymbol{\mu}_t)\,\boldsymbol{\epsilon}_{t-j}+(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t})\sum_{j=0}^{\infty}(\boldsymbol{\mu}_t\otimes\boldsymbol{B}_{j,t})\,\boldsymbol{\epsilon}_{t-j}\\ &\quad+(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t})\sum_{j=0}^{\infty}(\boldsymbol{B}_{j,t}\otimes\boldsymbol{B}_{j,t})\,\mathrm{vec}[\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-j}^\top-\boldsymbol{I}_d]\\ &\quad+(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t})\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}(\boldsymbol{B}_{j+r,t}\otimes\boldsymbol{B}_{j,t})\mathrm{vec}[\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-j-r}^\top]\\ &\quad+(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t})\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}(\boldsymbol{B}_{j,t}\otimes\boldsymbol{B}_{j+r,t})\mathrm{vec}[\boldsymbol{\epsilon}_{t-j-r}\boldsymbol{\epsilon}_{t-j}^\top].\end{aligned}$$

Consequently, we obtain

$$\begin{aligned}&\left\|\sum_{t=1}^{T}\boldsymbol{W}_{T,t}\left(\boldsymbol{x}_t\boldsymbol{x}_t^\top-E\left(\boldsymbol{x}_t\boldsymbol{x}_t^\top\right)\right)\right\| \le\ 2\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t})\sum_{j=0}^{\infty}(\boldsymbol{\mu}_t\otimes\boldsymbol{B}_{j,t})\,\boldsymbol{\epsilon}_{t-j}\right\|\\ &+\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t})\sum_{j=0}^{\infty}(\boldsymbol{B}_{j,t}\otimes\boldsymbol{B}_{j,t})\,\mathrm{vec}[\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-j}^\top-\boldsymbol{I}_d]\right\|\\ &+2\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t})\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}(\boldsymbol{B}_{j+r,t}\otimes\boldsymbol{B}_{j,t})\mathrm{vec}[\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-j-r}^\top]\right\| :=\ I_{T,5}+I_{T,6}+I_{T,7}.\end{aligned}$$

By the development of $\sum_{t=1}^{T}\boldsymbol{W}_{T,t}\left(\boldsymbol{x}_t-E(\boldsymbol{x}_t)\right)$, it is easy to know that $I_{T,5}$ is $O_P(\sqrt{d_T})$.

For $I_{T,6}$, by Lemma A.3, write

$$
\begin{aligned}
I_{T,6} \leq & \left\| \sum_{t=1}^{T} (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \, \mathbb{B}_t^0(1) \left(\mathrm{vec}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top) - \mathrm{vec}(\boldsymbol{I}_d) \right) \right\| \\
& + \left\| (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,1}) \, \widetilde{\mathbb{B}}_1^0(L) \mathrm{vec}(\boldsymbol{\epsilon}_0 \boldsymbol{\epsilon}_0^\top) \right\| + \left\| (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,T}) \, \widetilde{\mathbb{B}}_T^0(L) \mathrm{vec}(\boldsymbol{\epsilon}_T \boldsymbol{\epsilon}_T^\top) \right\| \\
& + \left\| \sum_{t=1}^{T-1} \left((\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t+1}) \, \widetilde{\mathbb{B}}_{t+1}^0(L) - (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \, \widetilde{\mathbb{B}}_t^0(L) \right) \mathrm{vec} \left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top \right) \right\| \\
:= & \, I_{T,61} + I_{T,62} + I_{T,63} + I_{T,64}.
\end{aligned}
$$

Let $\boldsymbol{Z}_t = \mathrm{vec}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top - \boldsymbol{I}_d)$ for notational simplicity. For $I_{T,61}$, write

$$
\begin{aligned}
& E \left\| \sum_{t=1}^{T} (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \, \mathbb{B}_t^0(1) \boldsymbol{Z}_t \right\|^2 \\
\leq & M \left(\max_{t \geq 1} \sum_{j=0}^{\infty} \|\boldsymbol{B}_{j,t}\|^2 \right)^2 \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}\|^2 E \|\boldsymbol{Z}_t\|^2 \leq M \max_{t \geq 1} \|\boldsymbol{W}_{T,t}\| \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}\| = O(d_T),
\end{aligned}
$$

which implies that $I_{T,61} = O_P(\sqrt{d_T})$. Similar to the proof of $\boldsymbol{I}_{T,2}$ and $\boldsymbol{I}_{T,3}$, we can prove that $I_{T,62}$ and $I_{T,63}$ are $O_P(d_T)$. For $I_{T,64}$, we have

$$
\begin{aligned}
I_{T,64} \leq & \left\| \sum_{t=1}^{T-1} (\boldsymbol{I}_d \otimes (\boldsymbol{W}_{T,t+1} - \boldsymbol{W}_{T,t})) \, \widetilde{\mathbb{B}}_{t+1}^0(L) \mathrm{vec} \left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top \right) \right\| \\
& + \left\| \sum_{t=1}^{T-1} (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \left(\widetilde{\mathbb{B}}_{t+1}^0(L) - \widetilde{\mathbb{B}}_t^0(L) \right) \mathrm{vec} \left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top \right) \right\|.
\end{aligned}
$$

Similar to the proof of $\boldsymbol{I}_{T,4}$, by Lemma A.3, we can prove that $I_{T,64}$ is $O_P(d_T)$. Then we can conclude that $I_{T,6} = O_P(\sqrt{d_T})$.

For $I_{T,7}$, using Lemma A.3, we have

$$
\begin{aligned}
I_{T,7} \leq & \left\| \sum_{t=1}^{T} (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \sum_{r=1}^{\infty} \mathbb{B}_t^r(1) \mathrm{vec} \left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_{t-r}^\top \right) \right\| + \left\| (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,1}) \sum_{r=1}^{\infty} \widetilde{\mathbb{B}}_1^r(L) \mathrm{vec} \left(\boldsymbol{\epsilon}_0 \boldsymbol{\epsilon}_{-r}^\top \right) \right\| \\
& + \left\| (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,T}) \sum_{r=1}^{\infty} \widetilde{\mathbb{B}}_T^r(L) \mathrm{vec} \left(\boldsymbol{\epsilon}_T \boldsymbol{\epsilon}_{T-r}^\top \right) \right\| \\
& + \left\| \sum_{t=1}^{T-1} \sum_{r=1}^{\infty} \left((\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t+1}) \widetilde{\mathbb{B}}_{t+1}^r(L) - (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \widetilde{\mathbb{B}}_t^r(L) \right) \mathrm{vec} \left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_{t-r}^\top \right) \right\| \\
:= & \, I_{T,71} + I_{T,72} + I_{T,73} + I_{T,74}.
\end{aligned}
$$

For $I_{T,71}$, by Lemma A.3, we further write

$$
\begin{aligned}
&E\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t})\sum_{r=1}^{\infty}\mathbb{B}_t^r(1)\text{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^\top\right)\right\|^2\\
&= E\,\text{tr}\left\{\sum_{t=1}^{T}\sum_{s=1}^{T}(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t})\sum_{r,k=1}^{\infty}\mathbb{B}_t^r(1)\text{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^\top\right)\text{vec}^\top\left(\boldsymbol{\epsilon}_s\boldsymbol{\epsilon}_{s-k}^\top\right)\mathbb{B}_s^{k,\top}(1)(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,s}^\top)\right\}\\
&\leq M\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}\|^2\sum_{r=1}^{\infty}\|\mathbb{B}_t^r(1)\|^2\leq M\left(\max_{t\geq 1}\sum_{r=1}^{\infty}\|\mathbb{B}_t^r(1)\|\right)^2\max_{t\geq 1}\|\boldsymbol{W}_{T,t}\|\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}\|=O(d_T).
\end{aligned}
$$

In addition, similar to the proof of $\boldsymbol{I}_{T,2}$ to $\boldsymbol{I}_{T,4}$, we can show that $I_{T,72}$ to $I_{T,74}$ are $O_P(d_T)$.

Combining the above results, we have proved the case of $p=0$.

Similar to the development of $p=0$, we can consider the case with $p\geq 1$ given p is a fixed number. The details are omitted due to similarity. The proof is now completed. □

Proof of Lemma 2.3.

(1). By Lemma A.3, we have

$$
\boldsymbol{x}_t=\boldsymbol{\mu}_t+\mathbb{B}_t(1)\boldsymbol{\epsilon}_t+\widetilde{\mathbb{B}}_t(L)\boldsymbol{\epsilon}_{t-1}-\widetilde{\mathbb{B}}_t(L)\boldsymbol{\epsilon}_t.
$$

We are then able to write

$$
\begin{aligned}
&\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}\boldsymbol{W}_{T,t}(\tau)(\boldsymbol{x}_t-E(\boldsymbol{x}_t))\right\|\\
&\leq \sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}\boldsymbol{W}_{T,t}(\tau)\mathbb{B}_t(1)\boldsymbol{\epsilon}_t\right\|+\sup_{\tau\in[a,b]}\left\|\boldsymbol{W}_{T,1}(\tau)\widetilde{\mathbb{B}}_1(L)\boldsymbol{\epsilon}_0\right\|+\sup_{\tau\in[a,b]}\left\|\boldsymbol{W}_{T,T}(\tau)\widetilde{\mathbb{B}}_T(L)\boldsymbol{\epsilon}_T\right\|\\
&\quad+\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T-1}\left(\boldsymbol{W}_{T,t+1}(\tau)\widetilde{\mathbb{B}}_{t+1}(L)-\boldsymbol{W}_{T,t}(\tau)\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\epsilon}_t\right\|:=I_{T,1}+I_{T,2}+I_{T,3}+I_{T,4},
\end{aligned}
$$

where the definitions of $I_{T,j}$ for $j=1,\ldots,4$ are obvious.

By Lemma A.4, we have $I_{T,1}=O_P\left(\sqrt{d_T\log T}\right)$. Also, it's easy to see that $I_{T,2}=O_P(d_T)$ and $I_{T,3}=O_P(d_T)$, because $E\|\widetilde{\mathbb{B}}_1(L)\boldsymbol{\epsilon}_0\|<\infty$ and $E\|\widetilde{\mathbb{B}}_T(L)\boldsymbol{\epsilon}_T\|<\infty$ in view of the fact that

$$
\|\widetilde{\mathbb{B}}_1(1)\|\leq\sum_{j=0}^{\infty}\|\widetilde{\boldsymbol{B}}_{j,1}\|<\infty\quad\text{and}\quad\|\widetilde{\mathbb{B}}_T(1)\|\leq\sum_{j=0}^{\infty}\|\widetilde{\boldsymbol{B}}_{j,T}\|<\infty
$$

by Lemma A.3. Thus, we need only to consider $I_{T,4}$ below. Note that

(1). $\sum_{t=1}^{T-1}\|\widetilde{\mathbb{B}}_{t+1}(1)-\widetilde{\mathbb{B}}_t(1)\|=O(1)$ by Lemma A.3;

(2). $T^{2/\delta}d_T\log T\to 0$ and $\sup_{\tau\in[a,b]}\sum_{t=1}^{T-1}\|\boldsymbol{W}_{T,t+1}(\tau)-\boldsymbol{W}_{T,t}(\tau)\|=O\left(d_T\right)$ by the conditions in the body of this lemma;

(3). $\max_{1\leq t\leq T-1}\|\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\epsilon}_t\|=O_P(T^{1/\delta})$ by $E\|\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\epsilon}_t\|^\delta<\infty$ and

$$
\max_{1\leq t\leq T-1}\|\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\epsilon}_t\|\leq\left(\sum_{t=1}^{T-1}\|\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\epsilon}_t\|^\delta\right)^{1/\delta}=O_P(T^{1/\delta}).
$$

Hence, write

$$
\begin{aligned}
&\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T-1}\left(\boldsymbol{W}_{T,t+1}(\tau)\widetilde{\mathbb{B}}_{t+1}(L)-\boldsymbol{W}_{T,t}(\tau)\widetilde{\mathbb{B}}_{t}(L)\right)\boldsymbol{\epsilon}_t\right\|\\
=&\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T-1}(\boldsymbol{W}_{T,t+1}(\tau)-\boldsymbol{W}_{T,t}(\tau))\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\epsilon}_t+\boldsymbol{W}_{T,t}(\tau)(\widetilde{\mathbb{B}}_{t+1}(L)-\widetilde{\mathbb{B}}_{t}(L))\boldsymbol{\epsilon}_t\right\|\\
\leq&\max_{1\leq t\leq T-1}\|\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\epsilon}_t\|\cdot\sup_{\tau\in[a,b]}\sum_{t=1}^{T-1}\|\boldsymbol{W}_{T,t+1}(\tau)-\boldsymbol{W}_{T,t}(\tau)\|\\
&+\sup_{\tau\in[a,b],1\leq t\leq T}\|\boldsymbol{W}_{T,t}(\tau)\|\cdot\sum_{t=1}^{T-1}\|(\widetilde{\mathbb{B}}_{t+1}(L)-\widetilde{\mathbb{B}}_{t}(L))\boldsymbol{\epsilon}_t\|\\
=&O_P(T^{1/\delta}\cdot d_T)+O_P(d_T)=o_P(\sqrt{d_T\log T}).
\end{aligned}
$$

The first result then follows.

(2). Below, we consider $p=0$ only. The cases with fixed $p\geq 1$ can be verified in a similar manner, so omitted.

$$
\begin{aligned}
&\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}\mathrm{vec}\left(\boldsymbol{W}_{T,t}(\tau)(\boldsymbol{x}_t\boldsymbol{x}_t^\top-E(\boldsymbol{x}_t\boldsymbol{x}_t^\top))\right)\right\|\leq 2\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(\tau))\sum_{j=0}^{\infty}(\boldsymbol{B}_{j,t}\otimes\boldsymbol{\mu}_t)\boldsymbol{\epsilon}_{t-j}\right\|\\
&+\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(\tau))\sum_{j=0}^{\infty}(\boldsymbol{B}_{j,t}\otimes\boldsymbol{B}_{j,t})\left(\mathrm{vec}\left(\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-j}^\top\right)-\mathrm{vec}\left(\boldsymbol{I}_d\right)\right)\right\|\\
&+2\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(\tau))\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}(\boldsymbol{B}_{j+r,t}\otimes\boldsymbol{B}_{j,t})\mathrm{vec}\left(\boldsymbol{\epsilon}_{t-j}\epsilon_{t-j-r}\right)\right\|:=2I_{T,1}+I_{T,2}+2I_{T,3},
\end{aligned}
$$

wherein $I_{T,1}=O_P(\sqrt{d_T\log T})$ by a proof similar to the first result of this lemma.

Consider $I_{T,2}$. Using Lemma A.3, write

$$
\begin{aligned}
I_{T,2}\leq&\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(\tau))\mathbb{B}_t^0(1)\left(\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\right)-\mathrm{vec}\left(\boldsymbol{I}_d\right)\right)\right\|\\
&+\sup_{\tau\in[a,b]}\left\|(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,1}(\tau))\widetilde{\mathbb{B}}_1^0(L)\mathrm{vec}\left(\boldsymbol{\epsilon}_0\boldsymbol{\epsilon}_0^\top\right)\right\|+\sup_{\tau\in[a,b]}\left\|(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,T}(\tau))\widetilde{\mathbb{B}}_T^0(L)\mathrm{vec}\left(\boldsymbol{\epsilon}_T\boldsymbol{\epsilon}_T^\top\right)\right\|\\
&+\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T-1}\left((\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t+1}(\tau))\widetilde{\mathbb{B}}_{t+1}^0(L)-(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(\tau))\widetilde{\mathbb{B}}_t^0(L)\right)\cdot\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\right)\right\|\\
:=&I_{T,21}+I_{T,22}+I_{T,23}+I_{T,24}.
\end{aligned}
$$

By Lemma A.5, we have $I_{T,21}=O_P\left(\sqrt{d_T\log T}\right)$. Also, $I_{T,22}=O_P(d_T)$ and $I_{T,23}=O_P(d_T)$, because

$\|\widetilde{\mathbb{B}}_1^0(1)\| < \infty$ and $\|\widetilde{\mathbb{B}}_T^0(1)\| < \infty$ by Lemma A.3. Similar to the proof of the first result, for $I_{T,24}$, we write

$$
\begin{aligned}
&\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T-1}\Big((\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t+1}(\tau))\widetilde{\mathbb{B}}_{t+1}^0(L)-(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(\tau))\widetilde{\mathbb{B}}_t^0(L)\Big)\operatorname{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\right)\right\|\\
\leq{}& \sqrt{d}\sup_{\tau\in[a,b],1\leq t\leq T}\|\boldsymbol{W}_{T,t+1}(\tau)\|\cdot\sum_{t=1}^{T-1}\left\|\Big(\widetilde{\mathbb{B}}_{t+1}^0(L)-\widetilde{\mathbb{B}}_t^0(L)\Big)\operatorname{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\right)\right\|\\
&+\sqrt{d}\max_t\left\|\widetilde{\mathbb{B}}_t^0(L)\operatorname{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\right)\right\|\cdot\sup_{\tau\in[a,b]}\sum_{t=1}^{T-1}\|\boldsymbol{W}_{T,t+1}(\tau)-\boldsymbol{W}_{T,t}(\tau)\|=o_P\left(\sqrt{d_T\log T}\right),
\end{aligned}
$$

where we have used the following facts:

$$
\begin{aligned}
&(1).\ T^{4/\delta}d_T\log T\to 0;\\
&(2).\ \max_{t\geq 1}\left\|\widetilde{\mathbb{B}}_t^0(L)\operatorname{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\right)\right\|=O_P(T^{2/\delta});\\
&(3).\ \sup_{\tau\in[a,b]}\sum_{t=1}^{T-1}\|\boldsymbol{W}_{T,t+1}(\tau)-\boldsymbol{W}_{T,t}(\tau)\|=O(d_T);\\
&(4).\ \sum_{t=1}^{T-1}\left\|\widetilde{\mathbb{B}}_{t+1}^0(1)-\widetilde{\mathbb{B}}_t^0(1)\right\|=O(1).
\end{aligned}
$$

Then we can conclude that $I_{T,24}=O_P\left(\sqrt{d_T\log T}\right)$.

We now consider $I_{T,3}$. Using Lemma A.3, we have

$$
\begin{aligned}
&\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(\tau))\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}(\boldsymbol{B}_{j+r,t}\otimes\boldsymbol{B}_{j,t})\operatorname{vec}\left(\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-j-r}\right)\right\|\\
\leq{}& \sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(\tau))\boldsymbol{\zeta}_t\boldsymbol{\epsilon}_t\right\|+\sup_{\tau\in[a,b]}\left\|(\boldsymbol{I}_d\otimes W_{T,1}(\tau))\sum_{r=1}^{\infty}\widetilde{\mathbb{B}}_1^r(L)\operatorname{vec}\left(\boldsymbol{\epsilon}_0\boldsymbol{\epsilon}_{-r}^\top\right)\right\|\\
&+\sup_{\tau\in[a,b]}\left\|(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,T}(\tau))\sum_{r=1}^{\infty}\widetilde{\mathbb{B}}_T^r(L)\operatorname{vec}\left(\boldsymbol{\epsilon}_T\boldsymbol{\epsilon}_{T-r}^\top\right)\right\|\\
&+\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\Big((\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t+1}(\tau))\widetilde{\mathbb{B}}_{t+1}^r(L)-(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(\tau))\widetilde{\mathbb{B}}_t^r(L)\Big)\operatorname{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^\top\right)\right\|\\
:={}& I_{T,31}+I_{T,32}+I_{T,33}+I_{T,34},
\end{aligned}
$$

where $\boldsymbol{\zeta}_t$ is defined in Lemma A.5.

By Lemma A.5, $I_{T,31}=O_P\left(\sqrt{d_T\log T}\right)$. Moreover, $I_{T,32}=O_P(d_T)$ and $I_{T,33}=O_P(d_T)$, because $\sum_{r=1}^{\infty}\|\widetilde{\mathbb{B}}_1^r(1)\|<\infty$ and $\sum_{r=1}^{\infty}\|\widetilde{\mathbb{B}}_T^r(1)\|<\infty$ by Lemma A.3. For $I_{T,34}$, we write

$$
\begin{aligned}
&\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\Big((\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t+1}(\tau))\widetilde{\mathbb{B}}_{t+1}^r(L)-(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(\tau))\widetilde{\mathbb{B}}_t^r(L)\Big)\operatorname{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^\top\right)\right\|\\
\leq{}& \sqrt{d}\sup_{\tau\in[a,b],1\leq t\leq T}\|\boldsymbol{W}_{T,t}(\tau)\|\cdot\sum_{t=1}^{T-1}\left\|\sum_{r=1}^{\infty}\Big(\widetilde{\mathbb{B}}_{t+1}^r(L)-\widetilde{\mathbb{B}}_t^r(L)\Big)\operatorname{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^\top\right)\right\|\\
&+\sqrt{d}\max_t\left\|\sum_{r=1}^{\infty}\widetilde{\mathbb{B}}_t^r(L)\operatorname{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^\top\right)\right\|\cdot\sup_{\tau\in[a,b]}\sum_{t=1}^{T-1}\|\boldsymbol{W}_{T,t+1}(\tau)-\boldsymbol{W}_{T,t}(\tau)\|=o_P(\sqrt{d_T\log T}),
\end{aligned}
$$

where we have used the following results:

$$
\begin{aligned}
&(1).\ \ T^{4/\delta} d_T \log T \to 0;\\
&(2).\ \ \max_t \left\| \sum_{r=1}^{\infty} \widetilde{\mathbb{B}}_t^r(L) \text{vec}\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_{t-r}^\top\right) \right\| = O_P(T^{2/\delta});\\
&(3).\ \ \sum_{t=1}^{T-1} \sum_{r=1}^{\infty} \|\widetilde{\mathbb{B}}_{t+1}^r(1) - \widetilde{\mathbb{B}}_t^r(1)\| = O(1);\\
&(4).\ \ \sup_{\tau\in[a,b]} \sum_{t=1}^{T-1} \|\boldsymbol{W}_{T,t+1}(\tau) - \boldsymbol{W}_{T,t}(\tau)\| = O(d_T).
\end{aligned}
$$

Based on the above development, the proof of the case with $p = 0$ is done. The proof is now completed.

□

Proof of Lemma 2.4.

Similar to the proof of Lemma 2.2, we have

$$
\frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \boldsymbol{W}_{T,t}(\boldsymbol{x}_t - E(\boldsymbol{x})) = \frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \boldsymbol{W}_{T,t} \mathbb{B}_t(1) \boldsymbol{\epsilon}_t + o_P(1)
$$

as $d_T = o(1)$.

Since

$$
\begin{aligned}
\text{Var}\left(\frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \boldsymbol{W}_{T,t} \mathbb{B}_t(1) \boldsymbol{\epsilon}_t\right) &= \frac{1}{d_T} \sum_{t=1}^{T} \boldsymbol{W}_{T,t} \mathbb{B}_t(1) \mathbb{B}_t^\top(1) \boldsymbol{W}_{T,t}^\top \\
&\to \boldsymbol{\Sigma}_{\boldsymbol{W}},
\end{aligned}
$$

we then use the Cramér-Wold device to prove its asymptotic normality. That is to show that for any conformable vector $\boldsymbol{l}$,

$$
\frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \boldsymbol{l}^\top \boldsymbol{W}_{T,t} \mathbb{B}_t(1) \boldsymbol{\epsilon}_t \to_D N\left(\boldsymbol{0}, \boldsymbol{l}^\top \boldsymbol{\Sigma}_{\boldsymbol{W}} \boldsymbol{l}\right).
$$

Let $\boldsymbol{Z}_t = \frac{1}{\sqrt{d_T}} \boldsymbol{l}^\top \boldsymbol{W}_{T,t} \mathbb{B}_t(1) \boldsymbol{\epsilon}_t$. By the law of large numbers for martingale differences and the assumption $E\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top | \mathcal{F}_{t-1}\right) = \boldsymbol{I}_d$ a.s., we have $\sum_{t=1}^{T} \boldsymbol{Z}_t^2(\tau) \to_P \boldsymbol{\Sigma}_{\boldsymbol{W}}$.

Furthermore, for any $\nu > 0$, by both Hölder's and Markov's inequalities, we have

$$
\begin{aligned}
&\sum_{t=1}^{T} E\left(\boldsymbol{Z}_t^2(\tau) I\left(|\boldsymbol{Z}_t(\tau)| > \nu\right)\right) \\
&\le \sum_{t=1}^{T} \frac{1}{d_T} \|\boldsymbol{W}_{T,t}\|^\delta \left(E\|\mathbb{B}_t(1)\boldsymbol{\epsilon}_t\|^\delta\right)^{2/\delta} \left(\frac{E\|\mathbb{B}_t(1)\boldsymbol{\epsilon}_t\|^\delta}{(d_T)^{\delta/2} \nu^\delta}\right)^{(\delta-2)/\delta} \\
&= O\left(d_T^{(\delta-2)/2}\right) = o(1)
\end{aligned}
$$

since $\sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}\| = O(1)$ and $\max_{t\ge 1} \|\boldsymbol{W}_{T,t}\| = O(d_T)$. By Lemma A.1, the proof is now completed.

□

Proof of Lemma 2.5.

Let $\boldsymbol{e}_t = \boldsymbol{x}_t - E(\boldsymbol{x})_t$ and $Z_T^* = \frac{1}{\sqrt{d_T}}\sum_{t=1}^{T}\boldsymbol{d}^\top \boldsymbol{W}_{T,t}\boldsymbol{e}_t\xi_t^*$ for any conformable unit vector $\boldsymbol{d}$. Then, it suffices to show that

$$Z_T^* \to_{D^*} N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{W}}\right).$$

In the following, we first show that

$$\text{Var}^*(Z_T^*(\tau))^2 = \boldsymbol{\Sigma}_{\boldsymbol{W}} + o_P(1)$$

and then prove its normality by blocking techniques.

Conditioning on the original sample, we have

$$\begin{aligned}
E^*(Z_T^*)^2 &= \frac{1}{d_T}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{d}^\top \boldsymbol{W}_{T,t}\boldsymbol{e}_t\boldsymbol{e}_s^\top \boldsymbol{W}_{T,s}^\top \boldsymbol{d}E^*(\xi_t^*\xi_s^*)\\
= \frac{1}{d_T}\sum_{t=1}^{T}\boldsymbol{d}^\top \boldsymbol{W}_{T,t}\boldsymbol{e}_t\boldsymbol{e}_t^\top \boldsymbol{W}_{T,t}^\top \boldsymbol{d} &+ \frac{1}{d_T}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i}\boldsymbol{d}^\top \boldsymbol{W}_{T,t}\boldsymbol{e}_t\boldsymbol{e}_{t+i}^\top \boldsymbol{W}_{T,t+i}^\top \boldsymbol{d}a(i/l)\\
&+\frac{1}{d_T}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i}\boldsymbol{d}^\top \boldsymbol{W}_{T,t+i}\boldsymbol{e}_{t+i}\boldsymbol{e}_t^\top \boldsymbol{W}_{T,t}^\top \boldsymbol{d}a(i/l). \qquad (A.2)
\end{aligned}$$

For the first term on the right hand side of (A.2), similar to the proof of Lemma 2.2, it is straightforward to obtain that

$$\frac{1}{d_T}\sum_{t=1}^{T}\boldsymbol{d}^\top \boldsymbol{W}_{T,t}\boldsymbol{e}_t\boldsymbol{e}_t^\top \boldsymbol{W}_{T,t}^\top \boldsymbol{d} = \frac{1}{d_T}\sum_{t=1}^{T}\boldsymbol{d}^\top \boldsymbol{W}_{T,t}E(\boldsymbol{e}_t\boldsymbol{e}_t^\top)\boldsymbol{W}_{T,t}^\top \boldsymbol{d} + o_P(1).$$

For the second and third terms on the right hand side of (A.2), as $a(i/l) = 0$ for $i > l$, we have

$$\begin{aligned}
&E\left\|\sum_{i=1}^{T-1}\frac{1}{d_T}\sum_{t=1}^{T-i}\boldsymbol{W}_{T,t}\left(\boldsymbol{e}_t\boldsymbol{e}_{t+i}^\top - E(\boldsymbol{e}_t\boldsymbol{e}_{t+i}^\top)\right)\boldsymbol{W}_{T,t+i}^\top a(i/l)\right\|\\
&\leq \sum_{i=1}^{T-1}a(i/l)E\left\|\frac{1}{d_T}\sum_{t=1}^{T-i}\boldsymbol{W}_{T,t}\left(\boldsymbol{e}_t\boldsymbol{e}_{t+i}^\top - E(\boldsymbol{e}_t\boldsymbol{e}_{t+i}^\top)\right)\boldsymbol{W}_{T,t+i}^\top\right\|\\
&= l\cdot\sqrt{d_T} = o(1)
\end{aligned}$$

as we have $E\left\|\frac{1}{d_T}\sum_{t=1}^{T-i}\boldsymbol{W}_{T,t}\left(\boldsymbol{e}_t\boldsymbol{e}_{t+i}^\top - E(\boldsymbol{e}_t\boldsymbol{e}_{t+i}^\top)\right)\boldsymbol{W}_{T,t+i}^\top\right\| = O(\sqrt{d_T})$ by using similar arguments to those used in the proof of Lemma 2.2.

We now need only to focus on $\frac{1}{d_T}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i}\boldsymbol{W}_{T,t}E(\boldsymbol{e}_t\boldsymbol{e}_{t+i}^\top)\boldsymbol{W}_{T,t+i}^\top a(i/l)$. Note that

$$\begin{aligned}
&\frac{1}{d_T}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i}\boldsymbol{W}_{T,t}E(\boldsymbol{e}_t\boldsymbol{e}_{t+i}^\top)\boldsymbol{W}_{T,t+i}^\top a(i/l),\\
=&\frac{1}{d_T}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i}\boldsymbol{W}_{T,t}E(\boldsymbol{e}_t\boldsymbol{e}_{t+i}^\top)\boldsymbol{W}_{T,t+i}^\top + \frac{1}{d_T}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i}\boldsymbol{W}_{T,t}E(\boldsymbol{e}_t\boldsymbol{e}_{t+i}^\top)\boldsymbol{W}_{T,t+i}^\top(a(i/l)-1). \qquad (A.3)
\end{aligned}$$

It is then sufficient to show that the second term of the above equation is $o(1)$ since

$$\frac{1}{d_T}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{W}_{T,t}E(\boldsymbol{e}_t\boldsymbol{e}_s^\top)\boldsymbol{W}_{T,s}^\top = \text{Var}\left(\frac{1}{\sqrt{d_T}}\sum_{t=1}^{T}\boldsymbol{W}_{T,t}\boldsymbol{e}_t\right) \to \boldsymbol{\Sigma}_{\boldsymbol{W}}$$

by the proof of Lemma 2.2.

Let s_T satisfy $\frac{1}{s_T} + \frac{s_T^2}{l} \to 0$. The second term of (A.3) is then bounded by

$$\begin{aligned}
&\left\| \frac{1}{d_T} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{W}_{T,t} E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top) \boldsymbol{W}_{T,t+i}^\top (a(i/l) - 1) \right\| \\
\leq & M \sum_{i=1}^{s_T} \max_t \left\| E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top) \right\| |a(i/l) - 1| + M \sum_{i=d_T+1}^{\infty} \max_t \left\| E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top) \right\| |a(i/l) - 1| \\
\leq & M \sum_{i=1}^{s_T} (1 - a(i/l)) + M \sum_{i=d_T+1}^{\infty} \max_t \left\| E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top) \right\| = o(1),
\end{aligned}$$

since $|\sum_{i=1}^{s_T}(1 - a(i/l))| \leq M \sum_{i=1}^{s_T} i/l \leq M s_T^2/l = o(1)$ by Lipschitz continuity of $a(\cdot)$ and

$$\sum_{i=s_T+1}^{\infty} \max_t \left\| E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top) \right\| = o(1) \text{ as } s_T \to \infty.$$

Conditioning on the original sample, we now use standard arguments for using a block technique to show the asymptotic normality. Now let $Z_T^*(\tau) = \sum_{j=1}^k X_{T,j}^*(\tau) + \sum_{j=1}^k Y_{T,j}^*(\tau)$, where

$$X_{T,j}^*(\tau) = \frac{1}{\sqrt{d_T}} \sum_{t=B_j+1}^{B_j+r_1} \boldsymbol{d}^\top \boldsymbol{W}_{T,t} \boldsymbol{e}_t \xi_t^*, \quad Y_{T,j}^*(\tau) = \frac{1}{\sqrt{d_T}} \sum_{t=B_j+r_1+1}^{B_j+r_1+r_2} \boldsymbol{d}^\top \boldsymbol{W}_{T,t} \boldsymbol{e}_t \xi_t^*,$$

with $B_j = (j-1)(r_1 + r_2)$ and $k = \lceil T/(r_1 + r_2) \rceil$.

Let $r_1 = r_1(T)$ and $r_2 = r_2(T)$ satisfying $k \cdot r_2 \cdot d_T \to 0$, $r_1 \cdot d_T + l/(r_1) \to 0$ and $r_2/r_1 + l/r_2 \to 0$. We first show that $\sum_{j=1}^k Y_{T,j}^*(\tau) = o_P(1)$. Since $r_1 > l$ for large enough T and the blocks $Y_{T,j}^*$ are mutual independent conditionally on the original data, then we have

$$\begin{aligned}
& EE^* \left(\sum_{j=1}^k Y_{T,j}^*(\tau) \right)^2 = E \left(\sum_{j=1}^k E^*(Y_{T,j}^*(\tau))^2 \right) \\
\leq & \frac{1}{d_T} \sum_{i=-r_2+1}^{r_2-1} a(i/l) \max_t \left\| E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top) \right\| \sum_{j=1}^k \sum_{t=B_j+r_1+1}^{B_j+r_1+r_2-i} \|\boldsymbol{W}_{T,t}\| \cdot \|\boldsymbol{W}_{T,t+i}\| \\
\leq & M \frac{1}{d_T} \max_{0 \leq i \leq r_2 - 1} \sum_{j=1}^k \sum_{t=B_j+r_1+1}^{B_j+r_1+r_2-i} \|\boldsymbol{W}_{T,t}\| \cdot \|\boldsymbol{W}_{T,t+i}\| \leq \ M k r_2 d_T = o(1).
\end{aligned}$$

We employ Lindeberg CLT to establish the asymptotic normality of $\sum_{j=1}^k X_{T,j}^*(\tau)$ as the blocks $X_{T,j}^*(\tau)$ are independent when $r_2 > l$ for large enough T. As discussed before, we have already shown that the asymptotic variance is equal to $\boldsymbol{\Sigma}_{\boldsymbol{W}}$. We then need to verify that for every $\nu > 0$,

$$\sum_{j=1}^k E^* \left(\frac{X_{T,j}^*(\tau)^2}{E^* \left(\sum_{j=1}^k X_{T,j}^*(\tau) \right)^2} I \left(\frac{X_{T,j}^*(\tau)^2}{E^* \left(\sum_{j=1}^k X_{T,j}^*(\tau) \right)^2} > \nu \right) \right) = o_P(1).$$

Conditioning on the original sample, $\{\boldsymbol{e}_t \xi_t^*\}$ is an L_δ-mixingale sequence. By Hölder's inequality, Chebyshev's inequality and Lemma 2 in Hansen (1991), we have

$$\sum_{j=1}^{k} E^* \left(\frac{X_{T,j}^*(\tau)^2}{E^*\left(\sum_{j=1}^k X_{T,j}^*(\tau)\right)^2} I \left(\frac{X_{T,j}^*(\tau)^2}{E^*\left(\sum_{j=1}^k X_{T,j}^*(\tau)\right)^2} > \nu \right) \right)$$

$$\leq \sum_{j=1}^{k} \left(E^* \left(\frac{X_{T,j}^*(\tau)^2}{E^*(\sum_{j=1}^k X_{T,j}^*(\tau))^2} \right)^{\frac{\delta}{2}} \right)^{\frac{2}{\delta}} \left(\frac{E^* \left(\frac{X_{T,j}^*(\tau)^2}{E^*(\sum_{j=1}^k X_{T,j}^*(\tau))^2} \right)^{\frac{\delta}{2}}}{\nu^{\frac{\delta}{2}}} \right)^{\frac{\delta-2}{\delta}}$$

$$= \nu^{\frac{2-\delta}{2}} \sum_{j=1}^{k} \frac{E^*(X_{T,j}^*(\tau))^\delta}{\left(E^*(\sum_{j=1}^k X_{T,j}^*(\tau))^2\right)^{\frac{\delta}{2}}} \leq \nu^{\frac{2-\delta}{2}} \sum_{j=1}^{k} \frac{d_T^{-\delta/2} M \left[\sum_{t=B_j+1}^{B_j+r_1} \left(\boldsymbol{d}^\top \boldsymbol{W}_{T,t} \boldsymbol{e}_t\right)^2\right]^{\delta/2}}{\left(E^*(\sum_{j=1}^k X_{T,j}^*(\tau))^2\right)^{\frac{\delta}{2}}}$$

$$\leq \nu^{\frac{2-\delta}{2}} \sum_{j=1}^{k} \frac{M d_T^{-\delta/2} r_1^{\delta/2-1} \sum_{t=B_j+1}^{B_j+r_1} \left(\boldsymbol{d}^\top \boldsymbol{W}_{T,t} \boldsymbol{e}_t\right)^\delta}{\left(E^*(\sum_{j=1}^k X_{T,j}^*(\tau))^2\right)^{\frac{\delta}{2}}}$$

$$\leq \nu^{\frac{2-\delta}{2}} \frac{M d_T^{-\delta/2} r_1^{\delta/2-1} d_T^{\delta-1} \sum_{t=1}^T \|\boldsymbol{W}_{T,t}\| \cdot \|\boldsymbol{e}_t\|^\delta}{\left(E^*(\sum_{j=1}^k X_{T,j}^*(\tau))^2\right)^{\frac{\delta}{2}}} = O_P((d_T r_1)^{\delta/2-1}) = o_P(1).$$

Combining the above results, we have

$$Z_T^* \to_{D^*} N\left(\boldsymbol{0}, \boldsymbol{\Sigma_W}\right).$$

The proof is now completed. □

Proof of Lemma 2.6.

Define $\boldsymbol{\Xi}_i = \frac{1}{d_T} \sum_{t=1}^{T-i} \boldsymbol{W}_{T,t} E\left(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top\right) \boldsymbol{W}_{T,t+i}^\top$ with $\boldsymbol{e}_t = \sum_{j=0}^\infty \boldsymbol{B}_{j,t} \boldsymbol{\epsilon}_t$. Write

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{W}} = \underbrace{\boldsymbol{\Xi}_0 + \sum_{i=1}^{b} \psi\left(i/b\right) \left(\boldsymbol{\Xi}_i + \boldsymbol{\Xi}_i^\top\right)}_{I_{T,1}} + \underbrace{\widehat{\boldsymbol{\Xi}}_0 - \boldsymbol{\Xi}_0}_{I_{T,2}}$$
$$+ \underbrace{\sum_{i=1}^{b} \psi\left(i/b\right) \left(\widehat{\boldsymbol{\Xi}}_i - \boldsymbol{\Xi}_i + \widehat{\boldsymbol{\Xi}}_i^\top - \boldsymbol{\Xi}_i^\top\right)}_{I_{T,3}}.$$

We next prove $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{W}} \to_P \boldsymbol{\Sigma_W}$ by showing that $I_{T,1} \to \boldsymbol{\Sigma_W}$, $I_{T,2} = o_P(1)$ and $I_{T,3} = o_P(1)$ one by one.

Consider $\sum_{i=1}^b \psi\left(i/b\right) \boldsymbol{\Xi}_i$. By the fact that $\max_t \sum_{j=1}^\infty j \left\|\boldsymbol{B}_{j,t}\right\| < \infty$, Lipschitz continuity of $\psi(\cdot)$, and $\psi(0) = 1$, we have

$$\left\|\sum_{i=1}^{b} (1 - \psi\left(i/b\right)) \boldsymbol{\Xi}_i\right\| \leq M \cdot \max_t \|\boldsymbol{W}_{T,t}\| \cdot \frac{1}{d_T} \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}\| \sum_{j=0}^{\infty} \|\boldsymbol{B}_{j,t}\| \sum_{i=1}^{b} \frac{i}{b} \|\boldsymbol{B}_{j+i,t+i}\|$$
$$= O(1/b) = o(1).$$

Hence, we have $I_{T,1} \to \boldsymbol{\Sigma_W}$.

For $I_{T,2}$ and $I_{T,3}$, since $\sum_{i=1}^{b}|\psi(i/b)| = O(b)$, $b\sqrt{d_T} \to 0$ and $E\|\widehat{\boldsymbol{\Xi}}_i - \boldsymbol{\Xi}_i\| = O(\sqrt{d_T})$ (using similar arguments to those used in the proofs of Lemma 2.2), we have

$$E\|I_{T,3}\| \leq 2\max_{1\leq i\leq b} E\|\widehat{\boldsymbol{\Xi}}_i - \boldsymbol{\Xi}_i\| \cdot \sum_{i=1}^{b}|\psi(i/b)| = O(b\sqrt{d_T}) = o(1).$$

The proof is now completed. □

Proof of Theorem 3.1.

Since $\frac{1}{T}\sum_{t=1}^{T} K_h(\tau_t - \tau) = 1 + O\left(\frac{1}{Th}\right)$, we have

$$\widehat{\boldsymbol{\mu}}(\tau) - E\left(\widehat{\boldsymbol{\mu}}(\tau)\right) = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_t - \boldsymbol{\mu}(\tau_t))K_h\left(\tau_t - \tau\right) + O_P\left(\frac{1}{Th}\right),$$

which follows that $\sqrt{Th}\left(\widehat{\boldsymbol{\mu}}(\tau) - E\left(\widehat{\boldsymbol{\mu}}(\tau)\right)\right) = \frac{1}{\sqrt{Th}}\sum_{t=1}^{T}(\boldsymbol{x}_t - \boldsymbol{\mu}(\tau_t))K\left(\frac{\tau_t-\tau}{h}\right) + o_P(1)$.

As $\frac{1}{T}\sum_{t=1}^{T} K_h(\tau_t-\tau) = O(1)$, $\max_t \frac{1}{T}K_h(\tau_t-\tau) = O(1/(Th))$ and $\frac{1}{T}\sum_{t=1}^{T-1}(K_h(\tau_{t+1}-\tau)-K_h(\tau_t-\tau)) = O(1/(Th))$, by Lemma 2.4, we have

$$\begin{aligned}
&\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}(\boldsymbol{x}_t - \boldsymbol{\mu}(\tau_t))K\left(\frac{\tau_t-\tau}{h}\right)\\
&= \frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\mathbb{B}_t(1)\boldsymbol{\epsilon}_t K\left(\frac{\tau_t-\tau}{h}\right) + \frac{1}{\sqrt{Th}}\widetilde{\mathbb{B}}_1(L)\boldsymbol{\epsilon}_0 K\left(\frac{\tau_1-\tau}{h}\right) - \frac{1}{\sqrt{Th}}\widetilde{\mathbb{B}}_T(L)\boldsymbol{\epsilon}_T K\left(\frac{\tau_T-\tau}{h}\right)\\
&\quad + \frac{1}{\sqrt{Th}}\sum_{t=1}^{T-1}\left(\widetilde{\mathbb{B}}_{t+1}(L)K\left(\frac{\tau_{t+1}-\tau}{h}\right) - \widetilde{\mathbb{B}}_t(L)K\left(\frac{\tau_t-\tau}{h}\right)\right)\boldsymbol{\epsilon}_t\\
&\to_D N\left(\boldsymbol{0}, \widetilde{v}_0\left\{\sum_{j=0}^{\infty}\boldsymbol{B}_j(\tau)\right\}\left\{\sum_{j=0}^{\infty}\boldsymbol{B}_j^\top(\tau)\right\}\right).
\end{aligned}$$

For the bias term, we have for any $\tau \in (0,1)$

$$\frac{1}{Th}\sum_{t=1}^{T}\boldsymbol{\mu}(\tau_t)K\left(\frac{\tau_t-\tau}{h}\right) = \boldsymbol{\mu}(\tau) + \frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\mu}^{(2)}(\tau) + o(h^2) + O\left(\frac{1}{Th}\right).$$

The proof is now completed. □

Proof of Theorem 3.2.

Note that $\boldsymbol{x}_t^* = \widetilde{\boldsymbol{\mu}}(\tau_t) + \boldsymbol{e}_t^*$, so we can write

$$\widehat{\boldsymbol{\mu}}^*(\tau) - \widetilde{\boldsymbol{\mu}}(\tau) = \left(\sum_{t=1}^{T}W_{T,t}(\tau)\widetilde{\boldsymbol{\mu}}(\tau_t) - \widetilde{\boldsymbol{\mu}}(\tau)\right) + \sum_{t=1}^{T}W_{T,t}(\tau)\boldsymbol{e}_t^* := \boldsymbol{I}_{T,1} + \boldsymbol{I}_{T,2},$$

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

We start our investigation from $\boldsymbol{I}_{T,1}$, and write

$$\boldsymbol{I}_{T,1} = \left(\frac{1}{Th}\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)\frac{1}{T\widetilde{h}}\sum_{s=1}^{T}\boldsymbol{\mu}(\tau_s)K\left(\frac{\tau_s-\tau_t}{\widetilde{h}}\right) - \frac{1}{T\widetilde{h}}\sum_{s=1}^{T}\boldsymbol{\mu}(\tau_s)K\left(\frac{\tau_s-\tau}{\widetilde{h}}\right)\right)$$

$$
+\frac{1}{\sqrt{T\widetilde{h}}}\left(\sum_{t=1}^{T} K\left(\frac{\tau_t-\tau}{h}\right)\boldsymbol{Z}_T(\tau_t)-\boldsymbol{Z}_T(\tau)\right)+O_P\left(\frac{1}{Th}\right)
$$

$$
:=\boldsymbol{I}_{T,11}+\boldsymbol{I}_{T,12}+O_P\left(\frac{1}{Th}\right),
$$

where the definitions of $\boldsymbol{I}_{T,11}$ and $\boldsymbol{I}_{T,12}$ should be obvious, $\boldsymbol{Z}_T(\tau)=\frac{1}{\sqrt{T\widetilde{h}}}\sum_{t=1}^{T}\boldsymbol{e}_t K\left(\frac{\tau_t-\tau}{\widetilde{h}}\right)$ and $\boldsymbol{e}_t=\sum_{j=0}^{\infty}\boldsymbol{B}_j(\tau_t)\boldsymbol{\epsilon}_{t-j}$. Similar to the development of Lemma 2.2, we can show that $\|\boldsymbol{I}_{T,12}\|=O_P((T\widetilde{h})^{-1/2})$, which, along with the conditions of Theorem 3.2, yields

$$
\sqrt{Th}\|\boldsymbol{I}_{T,12}\|=O_P((h/\widetilde{h})^{1/2})=o_P(1).
$$

For $\boldsymbol{I}_{T,11}$, by the definition of Riemann integral, we have

$$
\begin{aligned}
\boldsymbol{I}_{T,11}&=\int_{-1}^{1}K(u)\int_{-1}^{1}K(v)\left(\boldsymbol{\mu}(\tau+v\widetilde{h}+uh)-\boldsymbol{\mu}(\tau+v\widetilde{h})\right)dvdu+O\left(\frac{1}{Th}\right)\\
&=\frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\mu}^{(2)}(\tau)+O(h^2(h+\widetilde{h}))+O\left(\frac{1}{Th}\right).
\end{aligned}
$$

Thus, we need only to focus on $\boldsymbol{I}_{T,2}$ and then show that

$$
\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)\boldsymbol{e}_t^*\to_{D^*}N\left(\boldsymbol{0},\tilde{v}_0\left\{\sum_{j=0}^{\infty}\boldsymbol{B}_j(\tau)\right\}\left\{\sum_{j=0}^{\infty}\boldsymbol{B}_j^\top(\tau)\right\}\right).
$$

Using the Cramér-Wold device, this is enough to show for any conformable unit vector $\boldsymbol{d}$,

$$
\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)\boldsymbol{d}^\top\boldsymbol{e}_t^*\to_{D^*}N\left(\boldsymbol{0},\tilde{v}_0\boldsymbol{d}^\top\left\{\sum_{j=0}^{\infty}\boldsymbol{B}_j(\tau)\right\}\left\{\sum_{j=0}^{\infty}\boldsymbol{B}_j^\top(\tau)\right\}\boldsymbol{d}\right).
$$

For $\forall\tau\in[h+\widetilde{h},1-h-\widetilde{h}]$, we write

$$
\begin{aligned}
\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)\boldsymbol{d}^\top\widetilde{\boldsymbol{e}}_t\xi_t^*&=Z_T^*(\tau)+\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)\boldsymbol{d}^\top(\widetilde{\boldsymbol{e}}_t-\boldsymbol{e}_t)\xi_t^*\\
&=Z_T^*(\tau)+o_{P^*}(1),
\end{aligned}\tag{A.4}
$$

where $Z_T^*(\tau)=\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)\boldsymbol{d}^\top\boldsymbol{e}_t\xi_t^*$, and the second equality follows from

$$
\begin{aligned}
&EE^*\left\|\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)\boldsymbol{d}^\top(\widetilde{\boldsymbol{e}}_t-\boldsymbol{e}_t)\xi_t^*\right\|^2\\
\leq&\max_{\lfloor T(\tau-h)\rfloor\leq t\leq\lceil T(\tau+h)\rceil}E\left\|\widetilde{\boldsymbol{e}}_t-\boldsymbol{e}_t\right\|^2\left(\frac{1}{Th}\sum_{t=1}^{T}\sum_{s=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)K\left(\frac{\tau_s-\tau}{h}\right)E^*(\xi_t^*\xi_s^*)\right)\\
=&\,O\left(\widetilde{h}^4+1/(T\widetilde{h})\right)O(l)=o\left(1\right),
\end{aligned}
$$

where $EE^*[\cdot]$ stands for taking the expectation of the variables with respect to the bootstrap draws, and then taking the exception with respect to the original sample.

By Lemma 2.5, we already have

$$Z_T^*(\tau) \to_{D^*} N\left(0, \tilde{v}_0 \boldsymbol{d}^\top \left\{\sum_{j=0}^{\infty} \boldsymbol{B}_j(\tau)\right\} \left\{\sum_{j=0}^{\infty} \boldsymbol{B}_j^\top(\tau)\right\} \boldsymbol{d}\right).$$

Combining the above results, we have

$$\sqrt{Th}\left(\widehat{\boldsymbol{\mu}}^*(\tau) - \widetilde{\boldsymbol{\mu}}(\tau)\right) \to_{D^*} N\left(\boldsymbol{\mu}_b(\tau), \widetilde{v}_0 \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau)\right). \tag{A.5}$$

The proof is now completed.

□

Proof of Theorem 3.3.

Define $\boldsymbol{\Xi}_i(\tau) = \frac{1}{Th}\sum_{t=1}^{T-i} E\left(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top\right) K\left(\frac{\tau_t - \tau}{h}\right)$ with $\boldsymbol{e}_t = \sum_{j=0}^{\infty} \boldsymbol{B}_j(\tau_t)\boldsymbol{\epsilon}_t$. Write

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}(\tau) = &\underbrace{\boldsymbol{\Xi}_0(\tau) + \sum_{i=1}^{b} \psi(i/b)\left(\boldsymbol{\Xi}_i(\tau) + \boldsymbol{\Xi}_i^\top(\tau)\right)}_{I_{T,1}} + \underbrace{\widehat{\boldsymbol{\Xi}}_0(\tau) - \boldsymbol{\Xi}_0(\tau)}_{I_{T,2}} \\ &+ \underbrace{\sum_{i=1}^{b} \psi(i/b)\left(\widehat{\boldsymbol{\Xi}}_i(\tau) - \boldsymbol{\Xi}_i(\tau) + \widehat{\boldsymbol{\Xi}}_i^\top(\tau) - \boldsymbol{\Xi}_i^\top(\tau)\right)}_{I_{T,3}}.\end{aligned}$$

We next prove $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}(\tau) \to_P \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau)$ by showing that $I_{T,1} \to \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau)$, $I_{T,2} = o_P(1)$ and $I_{T,3} = o_P(1)$ one by one.

First, consider $I_{T,1}$. For $\boldsymbol{\Xi}_0(\tau)$, since $\sum_{j=0}^{\infty}\left\|\boldsymbol{B}_j(\tau)\boldsymbol{B}_j^\top(\tau)\right\| \le \left(\sum_{j=0}^{\infty}\|\boldsymbol{B}_j(\tau)\|\right)^2 < \infty$, $\mathbb{B}_0(\tau) := \sum_{j=0}^{\infty} \boldsymbol{B}_j(\tau)\boldsymbol{B}_j^\top(\tau)$ converges uniformly over $[0,1]$ and is second–order continuously differentiable with $\mathbb{B}^{(1)}(\tau) = \sum_{j=0}^{\infty}\left(\boldsymbol{B}_j^{(1)}(\tau)\boldsymbol{B}_j^\top(\tau) + \boldsymbol{B}_j(\tau)\boldsymbol{B}_j^{(1),\top}(\tau)\right)$ and

$$\mathbb{B}^{(2)}(\tau) = \sum_{j=0}^{\infty}\left(\boldsymbol{B}_j^{(2)}(\tau)\boldsymbol{B}_j^\top(\tau) + 2\boldsymbol{B}_j^{(1)}(\tau)\boldsymbol{B}_j^{(1),\top}(\tau) + \boldsymbol{B}_j(\tau)\boldsymbol{B}_j^{(2),\top}(\tau)\right).$$

Hence,

$$\begin{aligned}\boldsymbol{\Xi}_0(\tau) &= \frac{1}{Th}\sum_{t=1}^{T}\sum_{j=0}^{\infty} \boldsymbol{B}_j(\tau_t)\boldsymbol{B}_j^\top(\tau_t) K\left(\frac{\tau_t - \tau}{h}\right) = \int_{-1}^{1} \mathbb{B}(\tau + uh) K(u)\,\mathrm{d}u + O(1/(Th)) \\ &= \sum_{j=0}^{\infty} \boldsymbol{B}_j(\tau)\boldsymbol{B}_j^\top(\tau) + o(1).\end{aligned}$$

Consider $\sum_{i=1}^{b} \psi(i/b)\,\boldsymbol{\Xi}_i(\tau)$. By the fact that $\sup_{\tau\in[0,1]}\sum_{j=1}^{\infty} j\,\|\boldsymbol{B}_j(\tau)\| < \infty$, the Lipschitz continuity of $\psi(\cdot)$, and $\psi(0) = 1$, we have

$$\begin{aligned}\left\|\sum_{i=1}^{b}(1 - \psi(i/b))\,\boldsymbol{\Xi}_i(\tau)\right\| &\le M \cdot \left(\frac{1}{Th}\sum_{t=1}^{T}\sum_{j=0}^{\infty}\|\boldsymbol{B}_j(\tau_t)\|\sum_{i=1}^{b}\frac{i}{b}\|\boldsymbol{B}_{j+i}(\tau_{t+i})\| K\left(\frac{\tau_t - \tau}{h}\right)\right) \\ &= O(1/b) = o(1).\end{aligned}$$

Hence, $\sum_{i=1}^{b}\psi\left(i/b\right)\mathbf{\Xi}_i(\tau)=\sum_{i=1}^{b}\mathbf{\Xi}_i(\tau)+o(1)$.

Since

$$\sum_{i=1}^{b}\sum_{j=0}^{\infty}\left\|\boldsymbol{B}_j(\tau)\boldsymbol{B}_{j+i}^{\top}(\tau)\right\|\leq\left(\sum_{j=0}^{\infty}\|\boldsymbol{B}_j(\tau)\|\right)^2<\infty,$$

$\mathbb{B}_b(\tau):=\sum_{i=1}^{b}\sum_{j=0}^{\infty}\boldsymbol{B}_j(\tau)\boldsymbol{B}_{j+i}^{\top}(\tau)$ converges uniformly over $[0,1]$ and is second–order continuously differentiable with $\mathbb{B}_b^{(1)}(\tau)=\sum_{i=1}^{b}\sum_{j=0}^{\infty}\left(\boldsymbol{B}_j^{(1)}(\tau)\boldsymbol{B}_{j+i}^{\top}(\tau)+\boldsymbol{B}_j(\tau)\boldsymbol{B}_{j+i}^{(1),\top}(\tau)\right)$ and

$$\mathbb{B}_b^{(2)}(\tau)=\sum_{i=1}^{b}\sum_{j=0}^{\infty}\left(\boldsymbol{B}_j^{(2)}(\tau)\boldsymbol{B}_{j+i}^{\top}(\tau)+2\boldsymbol{B}_j^{(1)}(\tau)\boldsymbol{B}_{j+i}^{(1),\top}(\tau)+\boldsymbol{B}_j(\tau)\boldsymbol{B}_j^{(2),\top}(\tau)\right).$$

In addition, since

$$\begin{aligned}\left\|\mathbb{B}_b(\tau_t)-\sum_{i=1}^{b}\sum_{j=0}^{\infty}\boldsymbol{B}_j(\tau_t)\boldsymbol{B}_{j+i}^{\top}(\tau_{t+i})\right\|&\leq\sum_{i=1}^{b}\sum_{j=0}^{\infty}\|\boldsymbol{B}_j(\tau_t)\|\cdot\|\boldsymbol{B}_{j+i}(\tau_t)-\boldsymbol{B}_{j+i}(\tau_{t+i})\|\\&\leq M\sum_{j=0}^{\infty}\|\boldsymbol{B}_j(\tau_t)\|\cdot\sum_{i=1}^{\infty}\left\|\frac{i}{T}\boldsymbol{B}_i^{(1)}(\tau_t)\right\|=O(1/T),\end{aligned}$$

we have

$$\begin{aligned}\sum_{i=1}^{b}\mathbf{\Xi}_i(\tau)&=\sum_{i=1}^{b}\frac{1}{Th}\sum_{t=1}^{T-i}\sum_{j=0}^{\infty}\boldsymbol{B}_j(\tau_t)\boldsymbol{B}_{j+i}^{\top}(\tau_{t+i})K\left(\frac{\tau_t-\tau}{h}\right)\\&=\int_{-1}^{1}\mathbb{B}_b(\tau+uh)K\left(u\right)\mathrm{d}u+O(1/(Th))=\sum_{i=1}^{b}\sum_{j=0}^{\infty}\boldsymbol{B}_j(\tau)\boldsymbol{B}_{j+i}^{\top}(\tau)+o(1).\end{aligned}$$

Hence, as $b\to\infty$, $I_{T,1}\to\mathbf{\Sigma}_{\boldsymbol{\mu}}(\tau)$.

For $I_{T,2}$, by Lemma 2.2 and $\frac{1}{Th}\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)=1+O\left(1/(Th)\right)$, we have $I_{T,2}=o_P(1)$.

Next, consider $I_{T,3}$. Since $\sum_{i=1}^{b}|\psi(i/b)|=O(b)$ and $b\max\{h^4,1/\sqrt{Th}\}\to 0$, we have

$$E\left\|I_{T,3}\right\|\leq 2\max_{1\leq i\leq b}E\left\|\widehat{\mathbf{\Xi}}_i(\tau)-\mathbf{\Xi}_i(\tau)\right\|\cdot\sum_{i=1}^{b}|\psi(i/b)|=O(h^4+1/\sqrt{Th})\cdot b=o(1)$$

if

$$\max_{1\leq i\leq b}E\left\|\widehat{\mathbf{\Xi}}_i(\tau)-\mathbf{\Xi}_i(\tau)\right\|=O\left(h^4+1/\sqrt{Th}\right).\tag{A.6}$$

Thus, we need only to complete the proof by proving (A.6). Thus, we write

$$\begin{aligned}\max_{1\leq i\leq b}E\left\|\widehat{\mathbf{\Xi}}_i(\tau)-\mathbf{\Xi}_i(\tau)\right\|&\leq\max_{1\leq i\leq b}E\left\|\frac{1}{Th}\sum_{t=1}^{T-i}\left(\boldsymbol{e}_t\boldsymbol{e}_{t+i}^{\top}-E\left(\boldsymbol{e}_t\boldsymbol{e}_{t+i}^{\top}\right)\right)K\left(\frac{\tau_t-\tau}{h}\right)\right\|\\&\quad+\max_{1\leq i\leq b}E\left\|\frac{1}{Th}\sum_{t=1}^{T-i}\left(\widehat{\boldsymbol{e}}_t\widehat{\boldsymbol{e}}_{t+i}^{\top}-\boldsymbol{e}_t\boldsymbol{e}_{t+i}^{\top}\right)K\left(\frac{\tau_t-\tau}{h}\right)\right\|+O(1/(Th))\\&=I_{T,4}+I_{T,5}+O(1/(Th)),\end{aligned}$$

where the definitions of $I_{T,4}$ and $I_{T,5}$ should be obvious. By Lemma 2.2, we have $I_{T,4}=O(1/\sqrt{Th})$.

For $I_{T,5}$, write

$$
\begin{aligned}
& E\left\|\frac{1}{Th}\sum_{t=1}^{T-i}\left(\widehat{\boldsymbol{e}}_t\widehat{\boldsymbol{e}}_{t+i}^\top-\boldsymbol{e}_t\boldsymbol{e}_{t+i}^\top\right)K\left(\frac{\tau_t-\tau}{h}\right)\right\| \\
= & E\left\|\frac{1}{Th}\sum_{t=1}^{T-i}\left(\widehat{\boldsymbol{e}}_t-\boldsymbol{e}_t\right)\left(\widehat{\boldsymbol{e}}_{t+i}-\boldsymbol{e}_{t+i}\right)^\top K\left(\frac{\tau_t-\tau}{h}\right)\right\|+E\left\|\frac{1}{Th}\sum_{t=1}^{T-i}\left(\widehat{\boldsymbol{e}}_t-\boldsymbol{e}_t\right)\boldsymbol{e}_{t+i}^\top K\left(\frac{\tau_t-\tau}{h}\right)\right\| \\
& +E\left\|\frac{1}{Th}\sum_{t=1}^{T-i}\boldsymbol{e}_t\left(\widehat{\boldsymbol{e}}_{t+i}-\boldsymbol{e}_{t+i}\right)^\top K\left(\frac{\tau_t-\tau}{h}\right)\right\|=I_{T,51}+I_{T,52}+I_{T,53}.
\end{aligned}
$$

For $I_{T,51}$, by Cauchy–Schwarz inequality and Theorem 3.1,

$$
I_{T,51}\leq\max_{\lceil T(\tau-h)\rceil\leq t\leq\lceil T(\tau+h)\rceil}E\left\|\widehat{\boldsymbol{e}}_t-\boldsymbol{e}_t\right\|^2\frac{1}{Th}\sum_{t=1}^{T-i}K\left(\frac{\tau_t-\tau}{h}\right)=O(h^4+1/(Th)).
$$

For $I_{T,52}$, again using $\frac{1}{Th}\sum_{t=1}^{T-i}K\left(\frac{\tau_t-\tau}{h}\right)=1+O\left(1/(Th)\right)$, we have

$$
\begin{aligned}
I_{T,52}\leq & E\left\|\frac{1}{Th}\sum_{t=1}^{T-i}\boldsymbol{M}_{\boldsymbol{\mu}}(\tau_t)\boldsymbol{e}_{t+i}^\top K\left(\frac{\tau_t-\tau}{h}\right)\right\| \\
& +E\left\|\frac{1}{Th}\sum_{t=1}^{T-i}\left(\frac{1}{Th}\sum_{s=1}^{T}\boldsymbol{e}_sK\left(\frac{\tau_t-\tau_s}{h}\right)\right)\boldsymbol{e}_{t+i}^\top K\left(\frac{\tau_t-\tau}{h}\right)\right\|+O(1/(Th)) \\
= & I_{T,521}+I_{T,522},
\end{aligned}
$$

where $\boldsymbol{M}_{\boldsymbol{\mu}}(\tau)=\boldsymbol{\mu}(\tau)-\frac{1}{Th}\sum_{s=1}^{T}\boldsymbol{\mu}(\tau_s)K\left(\frac{\tau_s-\tau}{h}\right)$ is a twice-differential function matrix satisfying that $\boldsymbol{M}_{\boldsymbol{\mu}}(\tau)=O(h^2)$. Hence, by Lemma 2.2, we have $I_{T,521}=O(h^2/\sqrt{Th})$. For $I_{T,522}$, by Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
I_{T,522}\leq & \frac{1}{Th}\sum_{t=1}^{T-i}\left\{E\left\|\frac{1}{Th}\sum_{s=1}^{T}\boldsymbol{e}_sK\left(\frac{\tau_t-\tau_s}{h}\right)\right\|^2\right\}^{1/2}\{E\left\|\boldsymbol{e}_{t+i}\right\|^2\}^{1/2}K\left(\frac{\tau_t-\tau}{h}\right) \\
\leq & \max_t\left\{E\left\|\frac{1}{Th}\sum_{s=1}^{T}\boldsymbol{e}_sK\left(\frac{\tau_t-\tau_s}{h}\right)\right\|^2\right\}^{1/2}\frac{1}{Th}\sum_{t=1}^{T-i}\{E\left\|\boldsymbol{e}_{t+i}\right\|^2\}^{1/2}K\left(\frac{\tau_t-\tau}{h}\right) \\
= & O(1/\sqrt{Th}).
\end{aligned}
$$

Putting the above results together, the proof is now completed. □

Proof of Theorem 4.1.

(1). For notational simplicity, let $\boldsymbol{Z}_{\tau,t}$ be the transpose of the t^{th} row of $\boldsymbol{Z}_\tau$. Also, we define

$$
\begin{aligned}
\boldsymbol{S}_{T,k}(\tau) &= \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Z}_t\boldsymbol{Z}_t^\top\left(\frac{\tau_t-\tau}{h}\right)^kK_h(\tau_t-\tau)\text{ for }0\leq k\leq 3, \\
\boldsymbol{M}(\tau_t) &= \boldsymbol{\beta}(\tau_t)-\boldsymbol{\beta}(\tau)-\boldsymbol{\beta}^{(1)}(\tau)(\tau_t-\tau)-\frac{1}{2}\boldsymbol{\beta}^{(2)}(\tau)(\tau_t-\tau)^2, \\
\boldsymbol{S}_T(\tau) &= \begin{pmatrix}\boldsymbol{S}_{T,0}(\tau) & \boldsymbol{S}_{T,1}(\tau)\\ \boldsymbol{S}_{T,1}(\tau) & \boldsymbol{S}_{T,2}(\tau)\end{pmatrix}.
\end{aligned}
$$

Since

$$\boldsymbol{y}_t = \boldsymbol{Z}_t^\top \left(\boldsymbol{\beta}(\tau) + \boldsymbol{\beta}^{(1)}(\tau)(\tau_t - \tau) + \frac{1}{2}\boldsymbol{\beta}^{(2)}(\tau)(\tau_t - \tau)^2 + \boldsymbol{M}(\tau_t) \right) + \boldsymbol{\eta}_t,$$

we can write

$$\begin{aligned}
&\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \\
&= [\boldsymbol{I}_l, \boldsymbol{0}_l] \left(\frac{1}{T}\sum_{t=1}^T \boldsymbol{Z}_{\tau,t}\boldsymbol{Z}_{,\tau,t}^\top K_h(\tau_t - \tau) \right)^{-1} \frac{1}{T}\sum_{t=1}^T \boldsymbol{Z}_{\tau,t}\boldsymbol{y}_t K_h(\tau_t - \tau) - \boldsymbol{\beta}(\tau) \\
&= [\boldsymbol{I}_l, \boldsymbol{0}_l]\boldsymbol{S}_T^{-1}(\tau) \begin{bmatrix} \boldsymbol{S}_{T,2}(\tau) \\ \boldsymbol{S}_{T,3}(\tau) \end{bmatrix} \frac{1}{2}h^2\boldsymbol{\beta}^{(2)}(\tau) + [\boldsymbol{I}_l, \boldsymbol{0}_l]\boldsymbol{S}_T^{-1}(\tau)\cdot\frac{1}{T}\sum_{t=1}^T \boldsymbol{Z}_{\tau,t}\boldsymbol{Z}_t^\top \boldsymbol{M}(\tau_t)K_h(\tau_t - \tau) \\
&\quad + [\boldsymbol{I}_l, \boldsymbol{0}_l]\boldsymbol{S}_T^{-1}(\tau)\cdot\frac{1}{T}\sum_{t=1}^T \boldsymbol{Z}_{\tau,t}\boldsymbol{\eta}_t K_h(\tau_t - \tau) = I_{T,1} + I_{T,2} + I_{T,3},
\end{aligned}$$

where the definitions of $I_{T,1}$ to $I_{T,3}$ should be obvious.

By Lemma A.6 and Lemma 2.2, we have

$$\begin{aligned}
I_{T,1} &= \frac{1}{2}h^2\widetilde{c}_2\boldsymbol{\beta}^{(2)}(\tau) + O_P\left(h^2(h^2 + (Th)^{-1/2})\right), \\
I_{T,2} &= o_P(h^2).
\end{aligned}$$

Thus, we focus on $I_{T,3}$ below. For any $\tau \in (0,1)$, as $\{\boldsymbol{Z}_t\boldsymbol{\eta}_t\}$ is a sequence of martingale differences, by Lemma 2.2 and the martingale central limit theory, we have

$$\sqrt{Th}I_{T,3} = \left(\boldsymbol{\Sigma}_{\boldsymbol{z}}^{-1}(\tau)\otimes \boldsymbol{I}_d\right)\left(\frac{\sqrt{Th}}{T}\sum_{t=1}^T \boldsymbol{Z}_t\boldsymbol{\eta}_t K_h(\tau_t - \tau)\right) + o_P(1) \to_D N\left(\boldsymbol{0}, \widetilde{v}_0\boldsymbol{V}(\tau)\right).$$

The proof of the first result of this theorem is now completed.

(2). The uniform convergence rate for $\widehat{\boldsymbol{\beta}}(\tau)$ follows directly from Lemmas 2.3 and A.7.3.

(3). By Lemma 2.2, we have

$$\left\|\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{z}}(\tau) - \boldsymbol{\Sigma}_{\boldsymbol{z}}(\tau)\right\| = o_P(1).$$

Then we need only to focus on the rate associated with $\widehat{\boldsymbol{\Omega}}(\tau)$. For notational simplicity, we ignore the $\frac{1}{T}\sum_{t=1}^T K_h(\tau_t - \tau)$, because of $\frac{1}{T}\sum_{t=1}^T K_h(\tau_t - \tau) = 1 + O((Th)^{-1})$.

Write

$$\begin{aligned}
\widehat{\boldsymbol{\Omega}}(\tau) &= \frac{1}{Th}\sum_{t=1}^T \widehat{\boldsymbol{\eta}}_t\widehat{\boldsymbol{\eta}}_t^\top K\left(\frac{\tau_t - \tau}{h}\right) \\
&= \frac{1}{Th}\sum_{t=1}^T (\boldsymbol{\eta}_t + \widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)(\boldsymbol{\eta}_t + \widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)^\top K\left(\frac{\tau_t - \tau}{h}\right) \\
&= \frac{1}{Th}\sum_{t=1}^T \boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top K\left(\frac{\tau_t - \tau}{h}\right) + \frac{1}{Th}\sum_{t=1}^T (\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)(\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)^\top K\left(\frac{\tau_t - \tau}{h}\right)
\end{aligned}$$

$$+\frac{1}{Th}\sum_{t=1}^{T}\boldsymbol{\eta}_t(\widehat{\boldsymbol{\eta}}_t-\boldsymbol{\eta}_t)^\top K\left(\frac{\tau_t-\tau}{h}\right)+\frac{1}{Th}\sum_{t=1}^{T}(\widehat{\boldsymbol{\eta}}_t-\boldsymbol{\eta}_t)\boldsymbol{\eta}_t^\top K\left(\frac{\tau_t-\tau}{h}\right)$$
$$:= I_{T,1}+I_{T,2}+I_{T,3}+I_{T,4}.$$

Consider $I_{T,1}$. Since $\{\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top - E\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top\right)\}$ is a sequence of martingale differences, we have

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top - E\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top\right)\right]K_h(\tau_t-\tau)\right\| = o_P(1).$$

Next, consider $I_{T,2}$. By the second result of this theorem

$$\|I_{T,2}\| \le \sup_{\tau\in[0,1]}\|\widehat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}(\tau)\|^2\cdot\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{Z}_t\|^2 K_h(\tau_t-\tau) = o_P(1).$$

Similarly, for $I_{T,3}$ and $I_{T,4}$, we have

$$\|I_{T,3}\| \le \sup_{\tau\in[0,1]}\left\|\widehat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}(\tau)\right\|\cdot\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{Z}_t\boldsymbol{\eta}_t\| K_h(\tau_t-\tau) = o_P(1).$$

The proof is now completed. □

Proof of Theorem 4.2.

(1). By Lemma A.8,

$$\begin{aligned}&\sqrt{T}\left(\widehat{\boldsymbol{c}}-\boldsymbol{c}\right)\\
&=\left(\boldsymbol{X}_{\boldsymbol{C}}^\top(\boldsymbol{I}_{dT}-\boldsymbol{S})^\top(\boldsymbol{I}_{dT}-\boldsymbol{S})\boldsymbol{X}_{\boldsymbol{C}}\right)^{-1}\boldsymbol{X}_{\boldsymbol{C}}^\top(\boldsymbol{I}_{dT}-\boldsymbol{S})^\top(\boldsymbol{I}_{dT}-\boldsymbol{S})\left(\begin{bmatrix}\boldsymbol{X}_{\widetilde{\boldsymbol{C}},1}^\top\boldsymbol{\theta}(\tau_1)\\ \vdots\\ \boldsymbol{X}_{\widetilde{\boldsymbol{C}},T}^\top\boldsymbol{\theta}(\tau_T)\end{bmatrix}+\boldsymbol{\eta}\right)\\
&=\boldsymbol{\Sigma}^{-1}\frac{1}{\sqrt{T}}\boldsymbol{X}_{\boldsymbol{C}}^\top(\boldsymbol{I}_{dT}-\boldsymbol{S})^\top(\boldsymbol{I}_{dT}-\boldsymbol{S})\boldsymbol{\eta}+o_P(1).\end{aligned}$$

Hence, it suffices to show that

$$\frac{1}{\sqrt{T}}\boldsymbol{X}_{\boldsymbol{C}}^\top(\boldsymbol{I}_{dT}-\boldsymbol{S})^\top(\boldsymbol{I}_{dT}-\boldsymbol{S})\boldsymbol{\eta}\to_D N\left(\boldsymbol{0},\boldsymbol{\Delta}\right).$$

By using the same argument as in the proof of Lemma A.8,

$$\frac{1}{\sqrt{T}}\boldsymbol{X}_{\boldsymbol{C}}^\top(\boldsymbol{I}_{dT}-\boldsymbol{S})^\top(\boldsymbol{I}_{dT}-\boldsymbol{S})\boldsymbol{\eta}=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\boldsymbol{X}_{\boldsymbol{C},t}-\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}(\tau_t)\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{C}}}^{-1}(\tau_t)\boldsymbol{X}_{\widetilde{\boldsymbol{C}},t}\right)\boldsymbol{\eta}_t+o_P(1).$$

Since $\left\{\left(\boldsymbol{X}_{\boldsymbol{C},t}-\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}(\tau_t)\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{C}}}^{-1}(\tau_t)\boldsymbol{X}_{\widetilde{\boldsymbol{C}},t}\right)\boldsymbol{\eta}_t\right\}$ is a sequence of martingale differences, the result follows by the central limit theorem for martingale differences. Note that the convergence of conditional variance can be proved by Lemma 2.2.

(2). Let $\boldsymbol{\Theta}(\tau) = [\boldsymbol{\theta}(\tau)^\top, h\boldsymbol{\theta}^{(1)}(\tau)^\top]^\top$. Note that

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau) &= [\boldsymbol{I}_{l-s}, \boldsymbol{0}_{l-s}](\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^\top \boldsymbol{K}_\tau \boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau})^{-1} \boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^\top \boldsymbol{K}_\tau \left(\boldsymbol{y} - \boldsymbol{X}_{\boldsymbol{C}} \widehat{\boldsymbol{c}} - \boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau} \boldsymbol{\Theta}(\tau) \right) \\
&= [\boldsymbol{I}_{l-s}, \boldsymbol{0}_{l-s}](\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^\top \boldsymbol{K}_\tau \boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau})^{-1} \boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^\top \boldsymbol{K}_\tau \boldsymbol{X}_{\boldsymbol{C}} (\boldsymbol{c} - \widehat{\boldsymbol{c}}) \\
&\quad + [\boldsymbol{I}_{l-s}, \boldsymbol{0}_{l-s}](\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^\top \boldsymbol{K}_\tau \boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau})^{-1} \boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^\top \boldsymbol{K}_\tau \left(\begin{bmatrix} \boldsymbol{X}_{\widetilde{\boldsymbol{C}},1}^\top \boldsymbol{\theta}(\tau_1) \\ \vdots \\ \boldsymbol{X}_{\widetilde{\boldsymbol{C}},T}^\top \boldsymbol{\theta}(\tau_T) \end{bmatrix} - \boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau} \boldsymbol{\Theta}(\tau) \right) \\
&\quad + [\boldsymbol{I}_{l-s}, \boldsymbol{0}_{l-s}](\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^\top \boldsymbol{K}_\tau \boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau})^{-1} \boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^\top \boldsymbol{K}_\tau \boldsymbol{\eta} \\
&:= \boldsymbol{I}_{T,1} + \boldsymbol{I}_{T,2} + \boldsymbol{I}_{T,3}.
\end{aligned}$$

By part (1), we have $\boldsymbol{I}_{T,1} = O_P(T^{-1/2})$. By using standard arguments of the local linear kernel method, we have $\boldsymbol{I}_{T,2} = \frac{1}{2} h^2 \widetilde{c}_2 \boldsymbol{\theta}^{(2)}(\tau) + o_P(h^2)$. Then, it suffices to show that

$$[\boldsymbol{I}_{l-s}, \boldsymbol{0}_{l-s}] \frac{\sqrt{h}}{\sqrt{T}} \boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau}^\top \boldsymbol{K}_\tau \boldsymbol{\eta} \to_D N\left(\boldsymbol{0}, \tilde{v}_0 \widetilde{\boldsymbol{C}} \left(\boldsymbol{\Sigma}_{\boldsymbol{z}}(\tau) \otimes \boldsymbol{\Omega}(\tau)\right) \widetilde{\boldsymbol{C}}^\top \right).$$

The above result follows by the central limit theorem for martingale differences.

□

Appendix B

This appendix gives the proofs of the preliminary lemmas of Section A.1.

Proof of Lemma A.3.

(1). The first result follows from the standard BN decomposition (e.g., Phillips and Solo, 1992), so the details are omitted. (2). For the second decomposition, write

$$\begin{aligned}
(1-L)\widetilde{\mathbb{B}}_t^r(L) &= \sum_{j=0}^{\infty} \left(L^j \sum_{k=j+1}^{\infty} (\boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t}) - L^{j+1} \sum_{k=j+1}^{\infty} (\boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t}) \right) \\
&= \sum_{j=0}^{\infty} \left(L^{j+1} \sum_{k=j+2}^{\infty} (\boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t}) - L^{j+1} \sum_{k=j+1}^{\infty} (\boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t}) \right) + \sum_{k=1}^{\infty} (\boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t}) \\
&= -\sum_{j=0}^{\infty} L^{j+1} (\boldsymbol{B}_{j+1+r,t} \otimes \boldsymbol{B}_{j+1,t}) + \sum_{k=1}^{\infty} (\boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t}) \\
&= -\sum_{j=0}^{\infty} L^{j} (\boldsymbol{B}_{j+r,t} \otimes \boldsymbol{B}_{j,t}) + \sum_{k=0}^{\infty} (\boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t}) = \mathbb{B}_t^r(1) - \mathbb{B}_t^r(L).
\end{aligned}$$

(3). By Assumption 1,

$$\max_{t\geq 1} \sum_{j=0}^{\infty} \left\| \widetilde{\boldsymbol{B}}_{j,t} \right\| \leq \max_{t\geq 1} \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} \|\boldsymbol{B}_{k,t}\| = \max_{t\geq 1} \sum_{j=1}^{\infty} j \, \|\boldsymbol{B}_{j,t}\| < \infty.$$

(4). By Assumption 1,

$$\sum_{t=1}^{T-1} \|\widetilde{\mathbb{B}}_{t+1}(1) - \widetilde{\mathbb{B}}_t(1)\| \leq \sum_{t=1}^{T-1} \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} \|\boldsymbol{B}_{k,t+1} - \boldsymbol{B}_{k,t}\| = \sum_{t=1}^{T-1} \sum_{j=1}^{\infty} j \|\boldsymbol{B}_{j,t+1} - \boldsymbol{B}_{j,t}\| < \infty.$$

(5). By Assumption 1,

$$\max_{t\geq 1} \sum_{j=0}^{\infty} \left\|\widetilde{\boldsymbol{B}}_{j,t}^r\right\| \leq \max_{t\geq 1} \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} \|\boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t}\| = \max_{t\geq 1} \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} \|\boldsymbol{B}_{k+r,t}\| \cdot \|\boldsymbol{B}_{k,t}\|$$

$$= \max_{t\geq 1} \sum_{j=1}^{\infty} j \|\boldsymbol{B}_{j+r,t}\| \cdot \|\boldsymbol{B}_{j,t}\| \leq M \max_{t\geq 1} \sum_{j=1}^{\infty} j \|\boldsymbol{B}_{j,t}\| < \infty.$$

(6). Write

$$\max_{t\geq 1} \sum_{r=1}^{\infty} \left\|\widetilde{\mathbb{B}}_t^r(1)\right\| \leq \max_{t\geq 1} \sum_{r=1}^{\infty} \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} \|\boldsymbol{B}_{k+r,t}\| \cdot \|\boldsymbol{B}_{k,t}\|$$

$$= \max_{t\geq 1} \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} \|\boldsymbol{B}_{k,t}\| \cdot \left(\sum_{r=1}^{\infty} \|\boldsymbol{B}_{k+r,t}\|\right) \leq \max_{t\geq 1} \left(\sum_{r=1}^{\infty} \|\boldsymbol{B}_{r,t}\|\right) \cdot \left(\sum_{j=1}^{\infty} j \|\boldsymbol{B}_{j,t}\|\right) < \infty.$$

(7). Write

$$\sum_{t=1}^{T-1} \sum_{r=0}^{\infty} \sum_{j=0}^{\infty} \left\|\widetilde{\mathbb{B}}_{t+1}^r(1) - \widetilde{\mathbb{B}}_t^r(1)\right\| \leq \sum_{t=1}^{T-1} \sum_{r=0}^{\infty} \sum_{j=0}^{\infty} \left\| \sum_{k=j+1}^{\infty} \boldsymbol{B}_{k+r,t+1} \otimes \boldsymbol{B}_{k,t+1} - \boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t} \right\|$$

$$\leq \sum_{t=1}^{T-1} \sum_{r=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} (\|\boldsymbol{B}_{k+r,t+1} - \boldsymbol{B}_{k+r,t}\| \cdot \|\boldsymbol{B}_{k,t+1}\| + \|\boldsymbol{B}_{k+r,t}\| \cdot \|\boldsymbol{B}_{k,t+1} - \boldsymbol{B}_{k,t}\|)$$

$$= \sum_{t=1}^{T-1} \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} \left(\|\boldsymbol{B}_{k,t+1}\| \cdot \sum_{r=0}^{\infty} \|\boldsymbol{B}_{k+r,t+1} - \boldsymbol{B}_{k+r,t}\| + \|\boldsymbol{B}_{k,t+1} - \boldsymbol{B}_{k,t}\| \cdot \sum_{r=0}^{\infty} \|\boldsymbol{B}_{k+r,t}\| \right)$$

$$\leq \left(\sum_{t=1}^{T-1} \sum_{r=1}^{\infty} \|\boldsymbol{B}_{r,t+1} - \boldsymbol{B}_{r,t}\| \right) \cdot \left(\max_{t\geq 1} \sum_{k=1}^{\infty} k \|\boldsymbol{B}_{k,t}\| \right)$$

$$+ \left(\sum_{t=1}^{T-1} \sum_{k=1}^{\infty} k \|\boldsymbol{B}_{k,t+1} - \boldsymbol{B}_{k,t}\| \right) \cdot \left(\max_{t\geq 1} \sum_{r=1}^{\infty} \|\boldsymbol{B}_{r,t}\| \right) < \infty.$$

The proof is now completed. □

Proof of Lemma A.4.

In the following proof, we cover the interval $[a, b]$ by a finite number of subintervals $\{S_l\}$, which are centered at s_l with the length denoted by δ_T. Denoting the number of these intervals by N_T, then $N_T = O(\delta_T^{-1})$. In addition, let $\delta_T = O(T^{-1}\gamma_T)$ with $\gamma_T = \sqrt{d_T \log T}$.

Write

$$\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}\boldsymbol{W}_{T,t}(\tau)\mathbb{B}_t(1)\boldsymbol{\epsilon}_t\right\| \leq \max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}\boldsymbol{W}_{T,t}(s_l)\mathbb{B}_t(1)\boldsymbol{\epsilon}_t\right\|$$
$$+\max_{1\leq l\leq N_T}\sup_{\tau\in S_l}\left\|\sum_{t=1}^{T}\left(\boldsymbol{W}_{T,t}(\tau)-\boldsymbol{W}_{T,t}(s_l)\right)\mathbb{B}_t(1)\boldsymbol{\epsilon}_t\right\| := J_{T,1}+J_{T,2}.$$

For $J_{T,2}$, since $\boldsymbol{W}_{T,t}(\cdot)$ is Lipschitz continuous and $\max_{t\geq 1}\|\mathbb{B}_t(1)\|<\infty$ by Assumption 1, we have

$$E|J_{T,2}| \leq \sum_{t=1}^{T}\max_{1\leq l\leq N_T}\sup_{\tau\in S_l}\|\boldsymbol{W}_{T,t}(\tau)-\boldsymbol{W}_{T,t}(s_l)\|\,E\,\|\mathbb{B}_t(1)\boldsymbol{\epsilon}_t\|$$
$$\leq MT\delta_T\max_{t\geq 1}E\|\mathbb{B}_t(1)\boldsymbol{\epsilon}_t\| = O(\gamma_T).$$

For $J_{T,1}$, we apply the truncation method. Define $\boldsymbol{\epsilon}_t' = \boldsymbol{\epsilon}_t I(\|\boldsymbol{\epsilon}_t\|\leq T^{\frac{1}{\delta}})$ and $\boldsymbol{\epsilon}_t''=\boldsymbol{\epsilon}_t-\boldsymbol{\epsilon}_t'$, where δ is defined in Assumption 2, and $I(\cdot)$ is the indicator function. Write

$$J_{T,1} = \max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}\boldsymbol{W}_{T,t}(s_l)\mathbb{B}_t(1)\left(\boldsymbol{\epsilon}_t'+\boldsymbol{\epsilon}_t''-E\left(\boldsymbol{\epsilon}_t'+\boldsymbol{\epsilon}_t''|\mathcal{F}_{t-1}\right)\right)\right\|$$
$$\leq \max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}\boldsymbol{W}_{T,t}(s_l)\mathbb{B}_t(1)\left(\boldsymbol{\epsilon}_t'-E(\boldsymbol{\epsilon}_t'|\mathcal{F}_{t-1})\right)\right\| + \max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}\boldsymbol{W}_{T,t}(s_l)\mathbb{B}_t(1)\boldsymbol{\epsilon}_t''\right\|$$
$$+\max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}\boldsymbol{W}_{T,t}(s_l)\mathbb{B}_t(1)E(\boldsymbol{\epsilon}_t''|\mathcal{F}_{t-1})\right\| := J_{T,11}+J_{T,12}+J_{T,13}.$$

Start from $J_{T,12}$. By Hölder's inequality and Markov's inequality,

$$E|J_{T,12}| \leq O(1)d_T\sum_{t=1}^{T}E\left\|\boldsymbol{\epsilon}_t''\right\| = O(1)d_T\sum_{t=1}^{T}E\|\boldsymbol{\epsilon}_t I(\|\boldsymbol{\epsilon}_t\|\geq T^{\frac{1}{\delta}})\|$$
$$\leq O(1)d_T\sum_{t=1}^{T}\left\{E\|\boldsymbol{\epsilon}_t\|^{\delta}\right\}^{\frac{1}{\delta}}\left\{E\|I(\|\boldsymbol{\epsilon}_t\|\geq T^{\frac{1}{\delta}})\|\right\}^{\frac{\delta-1}{\delta}}$$
$$= O(1)d_T\sum_{t=1}^{T}\left\{E\|\boldsymbol{\epsilon}_t\|^{\delta}\right\}^{\frac{1}{\delta}}\left\{\Pr(\|\boldsymbol{\epsilon}_t\|\geq T^{\frac{1}{\delta}})\right\}^{\frac{\delta-1}{\delta}}$$
$$\leq O(1)d_T\sum_{t=1}^{T}\left\{E\|\boldsymbol{\epsilon}_t\|^{\delta}\right\}^{\frac{1}{\delta}}\left\{\frac{E\|\boldsymbol{\epsilon}_t\|^{\delta}}{T}\right\}^{\frac{\delta-1}{\delta}} = O(T^{\frac{1}{\delta}}d_T) = o\left(\sqrt{d_T\log T}\right),$$

where the second inequality follows from Hölder's inequality, and the third inequality follows from Markov's inequality. Similarly, $J_{T,13}=O_P(T^{\frac{1}{\delta}}d_T)=o_P\left(\sqrt{d_T\log T}\right)$.

We now turn to $J_{T,11}$. For notational simplicity, let $\boldsymbol{Y}_t=\boldsymbol{W}_{T,t}(s_l)\mathbb{B}_t(1)\left(\boldsymbol{\epsilon}_t'-E(\boldsymbol{\epsilon}_t'|\mathcal{F}_{t-1})\right)$ for $1\leq t\leq T$ and $A_T = 2T^{\frac{1}{\delta}}d_T\max_{t\geq 1}\|\mathbb{B}_t(1)\|$. Simple algebra shows that $\|\boldsymbol{Y}_t\|\leq A_T$ uniformly in t and s_l. By Assumption 2 and the first condition in the body of this lemma,

$$\max_{1\leq l\leq N_T}\sum_{t=1}^{T}E\left(\|\boldsymbol{Y}_t\|^2|\mathcal{F}_{t-1}\right) \leq Md_T\max_{1\leq l\leq N_T}\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}(s_l)\|\,E\left(\|\boldsymbol{\epsilon}_t\|^2|\mathcal{F}_{t-1}\right) = O_{a.s.}(d_T).$$

By Lemma A.2 and $T^{\frac{2}{\delta}} d_T \log T \to 0$, choose some $\beta > 0$ (such as $\beta = 4$), and write

$$
\begin{aligned}
&\Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T\right) \\
= &\Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T,\ \max_{1\le l\le N_T}\left\|\sum_{t=1}^T E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| \le M d_T\right) \\
&+\Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T,\ \max_{1\le l\le N_T}\left\|\sum_{t=1}^T E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| > M d_T\right) \\
\le &\Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T,\ \max_{1\le l\le N_T}\left\|\sum_{t=1}^T E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| \le M d_T\right) \\
&+\Pr\left(\max_{1\le l\le N_T}\left\|\sum_{t=1}^T E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| > M d_T\right) \\
\le &N_T \exp\left(-\frac{\beta M\gamma_T^2}{2(M d_T + \gamma_T 2A_T)}\right) \le N_T\exp\left(-\frac{\beta}{2}\log T\right) = O(\delta_T^{-1})T^{-\frac{\beta}{2}} \to 0.
\end{aligned}
$$

Based on the above development, the proof is now completed. □

Proof of Lemma A.5.

(1). Similar to the proof of Lemma A.4, we use a finite number of subintervals $\{S_l\}$ to cover the interval $[a,b]$, which are centered at s_l with the length δ_T. Denote the number of these intervals by N_T then $N_T = O(\delta_T^{-1})$. In addition, let $\delta_T = O(T^{-1}\gamma_T)$ with $\gamma_T = \sqrt{d_T\log T}$.

$$
\begin{aligned}
&\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^T(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t}(\tau))\mathbb{B}_t^0(1)\left(\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\right)-\mathrm{vec}(\boldsymbol{I}_d)\right)\right\| \\
\le &\max_{1\le l\le N_T}\left\|\sum_{t=1}^T(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t}(s_l))\mathbb{B}_t^0(1)\left(\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\right)-\mathrm{vec}(\boldsymbol{I}_d)\right)\right\| \\
&+\max_{1\le l\le N_T}\sup_{\tau\in S_l}\left\|\sum_{t=1}^T\left(\boldsymbol{I}_d\otimes(\boldsymbol{W}_{T,t}(\tau)-\boldsymbol{W}_{T,t}(s_l))\right)\mathbb{B}_t^0(1)\left(\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\right)-\mathrm{vec}(\boldsymbol{I}_d)\right)\right\| \\
:= &J_{T,1}+J_{T,2}.
\end{aligned}
$$

Start from $J_{T,2}$. Similar to the proof of Lemma A.4, since

$$
\left\|\mathbb{B}_t^0(1)\right\| \le \sum_{j=0}^\infty \|\boldsymbol{B}_{j,t}\|^2 \le \left(\sum_{j=0}^\infty \|\boldsymbol{B}_{j,t}\|\right)^2 < \infty
$$

by Assumption 1, we have

$$
E|J_{T,2}| \le MT\delta_T \max_{t\ge 1} E\left\|\mathbb{B}_t^0(1)\left(\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\right)-\mathrm{vec}(\boldsymbol{I}_d)\right)\right\| = O(\gamma_T).
$$

We then apply the truncation method. Define $\boldsymbol{u}_t = \mathbb{B}_t^0(1)\left(\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\right)-\mathrm{vec}\left(\boldsymbol{I}_d\right)\right)$, $\boldsymbol{u}_t' = \boldsymbol{u}_t I(\|\boldsymbol{u}_t\| \le T^{\frac{2}{\delta}})$ and $\boldsymbol{u}_t'' = \boldsymbol{u}_t - \boldsymbol{u}_t'$. For $J_{T,1}$, write

$$
J_{T,1} = \max_{1\le l\le N_T}\left\|\sum_{t=1}^T(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t}(s_l))\left(\boldsymbol{u}_t'+\boldsymbol{u}_t''-E(\boldsymbol{u}_t'+\boldsymbol{u}_t''|\mathcal{F}_{t-1})\right)\right\|
$$

$$\leq \max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t}(s_l))\left(\boldsymbol{u}_t'-E(\boldsymbol{u}_t'|\mathcal{F}_{t-1})\right)\right\|+\max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t}(s_l))\boldsymbol{u}_t''\right\|$$

$$+\max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t}(s_l))E(\boldsymbol{u}_t''|\mathcal{F}_{t-1})\right\| := J_{T,11}+J_{T,12}+J_{T,13}.$$

As in the proof of Lemma A.4, we can show that $J_{T,12}=o_P(\sqrt{d_T\log T})$ and $J_{T,13}=o_P(\sqrt{d_T\log T})$ respectively. We focus on $J_{T,11}$ below.

For any $1\leq l\leq N_T$, let $\boldsymbol{Y}_t=(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t}(s_l))(\boldsymbol{u}_t'-E(\boldsymbol{u}_t'|\mathcal{F}_{t-1}))$. We then have $E\left(\boldsymbol{Y}_t|\mathcal{F}_{t-1}\right)=0$ and $\|\boldsymbol{Y}_t\|\leq 2T^{2/\delta}d_T\max_t\left\|\mathbb{B}_t^0(1)\right\|$. Since $\max_{t\geq 1}E\left(\|\boldsymbol{\epsilon}_t\|^4|\mathcal{F}_{t-1}\right)<\infty$ a.s., we can write

$$\max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\|\leq Md_T\max_{t\geq 1}E\left(\|\boldsymbol{u}_t\|^2|\mathcal{F}_{t-1}\right)\max_{1\leq l\leq N_T}\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}(s_l)\|=O_{a.s.}\left(d_T\right).$$

Similar to Lemma A.2, choose $\beta>0$ (such as $\beta=4$). In view of the fact that $T^{\frac{4}{\delta}}d_T\log T\to 0$, we write

$$\begin{aligned}
\Pr\left(J_{T,11}>\sqrt{\beta M}\gamma_T\right) =\ & \Pr\left(J_{T,11}>\sqrt{\beta M}\gamma_T,\ \max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\|\leq Md_T\right)\\
&+\Pr\left(J_{T,11}>\sqrt{\beta M}\gamma_T,\ \max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\|>Md_T\right)\\
\leq\ & \Pr\left(J_{T,11}>\sqrt{\beta M}\gamma_T,\ \max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\|\leq Md_T\right)\\
&+\Pr\left(\max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\|>Md_T\right)\\
\leq\ & N_T\exp\left(-\frac{\beta M\gamma_T^2}{2(Md_T+M\gamma_T T^{\frac{2}{\delta}}d_T)}\right)\\
\leq\ & N_T\exp\left(-\frac{\beta}{2}\log T\right)=N_T T^{-\frac{\beta}{2}}=o(1).
\end{aligned}$$

The first result then follows.

(2). Let $\{S_l\}$ be a finite number of subintervals covering the interval $[a,b]$, which are centered at s_l with the length δ_T. Denote the number of these intervals by N_T then $N_T=O(\delta_T^{-1})$. In addition, let $\delta_T=O(T^{-1}\gamma_T)$ with $\gamma_T=\sqrt{d_T\log T}$. Then

$$\begin{aligned}
\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}\left(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t}(\tau)\right)\boldsymbol{\zeta}_t\boldsymbol{\epsilon}_t\right\| \leq\ & \max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}\left(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t}(s_l)\right)\boldsymbol{\zeta}_t\boldsymbol{\epsilon}_t\right\|\\
+\max_{1\leq l\leq N_T}\sup_{\tau\in S_l}&\left\|\sum_{t=1}^{T}\left(\boldsymbol{I}_d\otimes\left(\boldsymbol{W}_{T,t}(\tau)-\boldsymbol{W}_{T,t}(s_l)\right)\right)\boldsymbol{\zeta}_t\boldsymbol{\epsilon}_t\right\| :=\ J_{T,3}+J_{T,4}.
\end{aligned}$$

Consider $J_{T,4}$. By the fact that $|\mathrm{tr}(\boldsymbol{A})|\leq d\,\|\boldsymbol{A}\|$ for any $d\times d$ matrix $\boldsymbol{A}$ and Assumption 1,

$$E\left\|\boldsymbol{\zeta}_t\boldsymbol{\epsilon}_t\right\|=E\left\|\sum_{r=1}^{\infty}\left(\sum_{s=0}^{\infty}\boldsymbol{B}_{s+r,t}\otimes\boldsymbol{B}_{s,t}\right)\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^\top\right)\right\|$$

$$
\begin{aligned}
&\leq \left(E \left\| \sum_{r=1}^{\infty} \left(\sum_{s=0}^{\infty} \boldsymbol{B}_{s+r,t} \otimes \boldsymbol{B}_{s,t} \right) \text{vec} \left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_{t-r}^\top \right) \right\|^2 \right)^{1/2} \\
&\leq \left\{ \text{tr} \left[\sum_{r=1}^{\infty} \left(\sum_{s=0}^{\infty} \boldsymbol{B}_{s+r,t} \otimes \boldsymbol{B}_{s,t} \right) \cdot (\boldsymbol{I}_d \otimes \boldsymbol{I}_d) \cdot \left(\sum_{s=0}^{\infty} \boldsymbol{B}_{s+r,t} \otimes \boldsymbol{B}_{s,t} \right)^\top \right] \right\}^{1/2} \\
&\leq M \left(\sum_{r=1}^{\infty} \left\| \sum_{s=0}^{\infty} \boldsymbol{B}_{s+r,t} \otimes \boldsymbol{B}_{s,t} \right\|^2 \right)^{1/2} \leq M \left(\sum_{r=1}^{\infty} \left(\sum_{s=0}^{\infty} \|\boldsymbol{B}_{s+r,t}\|^2 \right) \cdot \left(\sum_{s=0}^{\infty} \|\boldsymbol{B}_{s,t}\|^2 \right) \right)^{1/2} \\
&= M \left(\left(\sum_{r=1}^{\infty} r \|\boldsymbol{B}_{r,t}\|^2 \right) \cdot \left(\sum_{s=0}^{\infty} \|\boldsymbol{B}_{s,t}\|^2 \right) \right)^{1/2} \leq M \left(\left(\sum_{r=1}^{\infty} r \|\boldsymbol{B}_{r,t}\| \right)^2 \cdot \left(\sum_{s=0}^{\infty} \|\boldsymbol{B}_{s,t}\|^2 \right) \right)^{1/2} < \infty.
\end{aligned}
$$

Similarly, we have $E|J_{T,4}| \leq MT\delta_T \max_{t\geq 1} E\,\|\boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t\| = O(\gamma_T)$.

Before investigating $J_{T,3}$, we first show that

$$
\max_{1\leq l\leq N_T} \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\|\, E\left(\|\boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t\|^2 \,|\mathcal{F}_{t-1} \right) = O_P(1). \tag{B.1}
$$

Note that

$$
\begin{aligned}
&\max_{1\leq l\leq N_T} \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\|\, E\left(\|\boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t\|^2 \,|\mathcal{F}_{t-1} \right) \\
\leq\ &\max_{1\leq l\leq N_T} \left| \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\| \left(E\left(\|\boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t\|^2 \,|\mathcal{F}_{t-1} \right) - E\,\|\boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t\|^2 \right) \right| \\
&+ \max_{1\leq l\leq N_T} \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\|\, E\,\|\boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t\|^2
\end{aligned}
$$

and $\max_{1\leq l\leq N_T} \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\|\, E\,\|\boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t\|^2 = O(1)$. Thus, to prove (B.1), it is sufficient to show

$$
\max_{1\leq l\leq N_T} \left| \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\| \left(E\left(\|\boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t\|^2 \,|\mathcal{F}_{t-1} \right) - E\,\|\boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t\|^2 \right) \right| = o_P(1).
$$

In order to do so, we write

$$
\begin{aligned}
&\max_{1\leq l\leq N_T} \left| \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\| \left(E\left(\|\boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t\|^2 \,|\mathcal{F}_{t-1} \right) - E\,\|\boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t\|^2 \right) \right| \\
=\ &\max_{1\leq l\leq N_T} \left| \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\| \,\text{tr} \left(\sum_{r,r^*=1}^{\infty} \mathbb{B}_t^r(1)(\boldsymbol{\epsilon}_{t-r}\boldsymbol{\epsilon}_{t-r^*}^\top \otimes \boldsymbol{I}_d)\mathbb{B}_t^{r^*,\top}(1) - \sum_{r=1}^{\infty} \mathbb{B}_t^r(1)\mathbb{B}_t^{r,\top}(1) \right) \right| \\
\leq\ & d^2 \cdot \max_{1\leq l\leq N_T} \left\| \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\| \sum_{r=1}^{\infty} (\mathbb{B}_t^r(1) \otimes \mathbb{B}_t^r(1)) \left(\text{vec}\left(\boldsymbol{\epsilon}_{t-r}\boldsymbol{\epsilon}_{t-r}^\top \otimes \boldsymbol{I}_d \right) - \text{vec}(\boldsymbol{I}_{d^2}) \right) \right\| \\
&+ 2d^2 \cdot \max_{1\leq l\leq N_T} \left\| \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\| \sum_{r=1}^{\infty}\sum_{j=1}^{\infty} \left(\mathbb{B}_t^{r+j}(1) \otimes \mathbb{B}_t^r(1) \right) \text{vec}\left(\boldsymbol{\epsilon}_{t-r}\boldsymbol{\epsilon}_{t-r-j}^\top \otimes \boldsymbol{I}_d \right) \right\| := J_{T,5} + J_{T,6}.
\end{aligned}
$$

Let $\mathbb{F}_{r,t}(L)=\sum_{j=1}^{\infty}\mathbb{B}_t^{r+j}(1)\otimes\mathbb{B}_t^{j}(1)L^j$. Similar to the second result of Lemma A.3, we have

$$\mathbb{F}_{r,t}(L)=\mathbb{F}_{r,t}(1)-(1-L)\widetilde{\mathbb{F}}_{r,t}(L) \tag{B.2}$$

where $\widetilde{\mathbb{F}}_{r,t}(L)=\sum_{j=1}^{\infty}\widetilde{\mathbb{F}}_{rj,t}L^j$ and $\widetilde{\mathbb{F}}_{rj,t}=\sum_{k=j+1}^{\infty}\mathbb{B}_t^{r+k}(1)\otimes\mathbb{B}_t^{k}(1)$. For notational simplicity, denote

$$\begin{aligned}\widetilde{X}_{at}&=\sum_{j=1}^{\infty}\left(\mathbb{B}_t^{j}(1)\otimes\mathbb{B}_t^{j}(1)\right)\mathrm{vec}\left(\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-j}^\top\otimes\boldsymbol{I}_d\right),\\ \widetilde{X}_{bt}&=\sum_{r=1}^{\infty}\sum_{j=1}^{\infty}\left(\mathbb{B}_t^{r+j}(1)\otimes\mathbb{B}_t^{j}(1)\right)\mathrm{vec}\left(\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-r-j}^\top\otimes\boldsymbol{I}_d\right).\end{aligned}$$

Applying (B.2) to $\widetilde{X}_{at}$ and $\widetilde{X}_{bt}$ yields that

$$\begin{aligned}\widetilde{X}_{at}&=\mathbb{F}_{0,t}(1)\mathrm{vec}\left(\boldsymbol{\epsilon}_{t}\boldsymbol{\epsilon}_{t}^\top\otimes\boldsymbol{I}_d\right)-(1-L)\widetilde{\mathbb{F}}_{0,t}(L)\mathrm{vec}\left(\boldsymbol{\epsilon}_{t}\boldsymbol{\epsilon}_{t}^\top\otimes\boldsymbol{I}_d\right),\\ \widetilde{X}_{bt}&=\sum_{r=1}^{\infty}\mathbb{F}_{r,t}(1)\mathrm{vec}\left(\boldsymbol{\epsilon}_{t}\boldsymbol{\epsilon}_{t-r}^\top\otimes\boldsymbol{I}_d\right)-(1-L)\sum_{r=1}^{\infty}\widetilde{\mathbb{F}}_{r,t}(L)\mathrm{vec}\left(\boldsymbol{\epsilon}_{t}\boldsymbol{\epsilon}_{t-r}^\top\otimes\boldsymbol{I}_d\right).\end{aligned}$$

For $J_{T,5}$, summing up $\widetilde{X}_{at}$ over t yields

$$\begin{aligned}&\max_{1\le l\le N_T}\left\|\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}(s_l)\|\sum_{r=1}^{\infty}\left(\mathbb{B}_t^{r}(1)\otimes\mathbb{B}_t^{r}(1)\right)\left(\mathrm{vec}\left(\boldsymbol{\epsilon}_{t-r}\boldsymbol{\epsilon}_{t-r}^\top\otimes\boldsymbol{I}_d\right)-\mathrm{vec}(\boldsymbol{I}_{d^2})\right)\right\|\\ \le\ &\max_{1\le l\le N_T}\left\|\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}(s_l)\|\,\mathbb{F}_{0,t}(1)\left(\mathrm{vec}\left(\boldsymbol{\epsilon}_{t}\boldsymbol{\epsilon}_{t}^\top\otimes\boldsymbol{I}_d\right)-\mathrm{vec}\left(\boldsymbol{I}_{d^2}\right)\right)\right\|\\ &+\max_{1\le l\le N_T}\left\|\|\boldsymbol{W}_{T,1}(s_l)\|\,\widetilde{\mathbb{F}}_{0,1}(L)\mathrm{vec}\left(\boldsymbol{\epsilon}_{0}\boldsymbol{\epsilon}_{0}^\top\otimes\boldsymbol{I}_d\right)\right\|+\sup_{0\le\tau\le1}\left\|\|\boldsymbol{W}_{T,T}(s_l)\|\,\widetilde{\mathbb{F}}_{0,T}(L)\mathrm{vec}\left(\boldsymbol{\epsilon}_{T}\boldsymbol{\epsilon}_{T}^\top\otimes\boldsymbol{I}_d\right)\right\|\\ &+\max_{1\le l\le N_T}\left\|\sum_{t=1}^{T-1}\left(\|\boldsymbol{W}_{T,t+1}(s_l)\|\,\widetilde{\mathbb{F}}_{0,t+1}(L)-\|\boldsymbol{W}_{T,t}(s_l)\|\,\widetilde{\mathbb{F}}_{0,t}(L)\right)\mathrm{vec}\left(\boldsymbol{\epsilon}_{t}\boldsymbol{\epsilon}_{t}^\top\otimes\boldsymbol{I}_d\right)\right\|\\ :=\ &J_{T,51}+J_{T,52}+J_{T,53}+J_{T,54}.\end{aligned}$$

Similar to the proof of Lemma A.5.1, we can show that $J_{T,51}=O_P\left(\sqrt{d_T\log T}\right)$, since

$$\begin{aligned}\max_{t\ge1}\|\mathbb{F}_{0,t}(1)\|&\le\max_{t\ge1}\sum_{j=1}^{\infty}\left\|\sum_{k=0}^{\infty}\boldsymbol{B}_{k+j,t}\otimes\boldsymbol{B}_{k,t}\right\|^2\le\max_{t\ge1}\sum_{j=1}^{\infty}\left(\sum_{k=0}^{\infty}\|\boldsymbol{B}_{k+j,t}\|^2\right)\left(\sum_{k=0}^{\infty}\|\boldsymbol{B}_{k,t}\|^2\right)\\ &\le\max_{t\ge1}\left(\sum_{k=0}^{\infty}\|\boldsymbol{B}_{k,t}\|^2\right)\left(\sum_{j=1}^{\infty}j\,\|\boldsymbol{B}_{j,t}\|^2\right)<\infty.\end{aligned}$$

Also, we can show that $J_{T,52}=O_P(d_T)$ and $J_{T,53}=O_P(d_T)$, since

$$\begin{aligned}\max_{t\ge1}\left\|\widetilde{\mathbb{F}}_{0,t}(1)\right\|&\le\max_{t\ge1}\sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}\left\|\mathbb{B}_t^{k}(1)\right\|^2\le\max_{t\ge1}\sum_{r=1}^{\infty}\sum_{k=r+1}^{\infty}\left(\sum_{j=0}^{\infty}\|\boldsymbol{B}_{j+k,t}\|^2\right)\left(\sum_{j=0}^{\infty}\|\boldsymbol{B}_{j,t}\|^2\right)\\ &\le\max_{t\ge1}\left(\sum_{j=0}^{\infty}\|\boldsymbol{B}_{j,t}\|^2\right)\left(\sum_{r=1}^{\infty}\sum_{k=r+1}^{\infty}(k-r)\,\|\boldsymbol{B}_{k,t}\|^2\right)\end{aligned}$$

$$\leq \max_{t\geq 1}\left(\sum_{j=0}^{\infty}\|\boldsymbol{B}_{j,t}\|^2\right)\left(\sum_{r=1}^{\infty}\frac{r(r+1)}{2}\|\boldsymbol{B}_{r+1,t}\|^2\right)$$
$$\leq \max_{t\geq 1}\left(\sum_{j=0}^{\infty}\|\boldsymbol{B}_{j,t}\|^2\right)\left(\sum_{j=1}^{\infty}j^2\|\boldsymbol{B}_{j,t}\|^2\right)<\infty.$$

We can easily show $J_{T,54}=o_P(1)$, since

$$\sup_{\tau\in[a,b]}\left(\sum_{t=1}^{T-1}\|\boldsymbol{W}_{T,t+1}(\tau)\|-\|\boldsymbol{W}_{T,t}(\tau)\|\right)\leq \sup_{\tau\in[a,b]}\sum_{t=1}^{T-1}\|\boldsymbol{W}_{T,t+1}(\tau)-\boldsymbol{W}_{T,t}(\tau)\|=o(1)$$

and

$$\sum_{t=1}^{T-1}\left\|\widetilde{\mathbb{F}}_{0,t+1}(1)-\widetilde{\mathbb{F}}_{0,t}(1)\right\|\leq \sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\sum_{k=r+1}^{\infty}\left\|\mathbb{B}_{t+1}^k(1)\otimes\mathbb{B}_{t+1}^k(1)-\mathbb{B}_t^k(1)\otimes\mathbb{B}_t^k(1)\right\|$$
$$\leq \sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\sum_{k=r+1}^{\infty}\left\|\mathbb{B}_{t+1}^k(1)-\mathbb{B}_t^k(1)\right\|\cdot\left(\left\|\mathbb{B}_{t+1}^k(1)\right\|+\left\|\mathbb{B}_t^k(1)\right\|\right)$$
$$\leq M\sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\sum_{k=r+1}^{\infty}\sum_{j=0}^{\infty}\|\boldsymbol{B}_{j+k,t+1}\otimes\boldsymbol{B}_{j,t+1}-\boldsymbol{B}_{j+k,t}\otimes\boldsymbol{B}_{j,t}\|$$
$$\leq M\sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\sum_{k=r+1}^{\infty}\sum_{j=0}^{\infty}(\|\boldsymbol{B}_{j+k,t+1}-\boldsymbol{B}_{j+k,t}\|\cdot\|\boldsymbol{B}_{j,t+1}\|+\|\boldsymbol{B}_{j,t+1}-\boldsymbol{B}_{j,t}\|\cdot\|\boldsymbol{B}_{j+k,t}\|)$$
$$\leq M\sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\sum_{k=r+1}^{\infty}\left(\|\boldsymbol{B}_{k,t+1}-\boldsymbol{B}_{k,t}\|\cdot\sum_{j=0}^{\infty}\|\boldsymbol{B}_{j,t+1}\|+\|\boldsymbol{B}_{k,t}\|\cdot\sum_{j=0}^{\infty}\|\boldsymbol{B}_{j,t+1}-\boldsymbol{B}_{j,t}\|\right)$$
$$\leq M\left(\sum_{t=1}^{T-1}\sum_{k=1}^{\infty}k\|\boldsymbol{B}_{k,t+1}-\boldsymbol{B}_{k,t}\|\right)\cdot\left(\max_{t\geq 1}\sum_{j=0}^{\infty}\|\boldsymbol{B}_{j,t+1}\|\right)$$
$$+M\left(\max_{t\geq 1}\sum_{k=1}^{\infty}k\|\boldsymbol{B}_{k,t}\|\right)\cdot\left(\sum_{t=1}^{T-1}\sum_{j=0}^{\infty}\|\boldsymbol{B}_{j,t+1}-\boldsymbol{B}_{j,t}\|\right)=O(1).$$

Based on the above development, we conclude that $J_{T,5}=o_P(1)$. Next, we focus on $J_{T,6}$, and write

$$\max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}(s_l)\|\sum_{r=1}^{\infty}\sum_{j=1}^{\infty}\mathbb{B}_t^{r+j}(1)\otimes\mathbb{B}_t^r(1)\mathrm{vec}\left(\boldsymbol{\epsilon}_{t-r}\boldsymbol{\epsilon}_{t-r-j}^\top\otimes\boldsymbol{I}_d\right)\right\|$$
$$\leq \max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}(s_l)\|\sum_{r=1}^{\infty}\mathbb{F}_{r,t}(1)\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^\top\otimes\boldsymbol{I}_d\right)\right\|$$
$$+\max_{1\leq l\leq N_T}\left\|\|\boldsymbol{W}_{T,1}(s_l)\|\sum_{r=1}^{\infty}\widetilde{\mathbb{F}}_{r,1}(L)\mathrm{vec}\left(\boldsymbol{\epsilon}_0\boldsymbol{\epsilon}_{-r}^\top\otimes\boldsymbol{I}_d\right)\right\|$$
$$+\max_{1\leq l\leq N_T}\left\|\|\boldsymbol{W}_{T,T}(s_l)\|\sum_{r=1}^{\infty}\widetilde{\mathbb{F}}_{r,T}(L)\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{T-r}^\top\otimes\boldsymbol{I}_d\right)\right\|$$
$$+\max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\left(\|\boldsymbol{W}_{T,t+1}(s_l)\|\widetilde{\mathbb{F}}_{r,t+1}(L)-\|\boldsymbol{W}_{T,t}(s_l)\|\widetilde{\mathbb{F}}_{r,t}(L)\right)\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^\top\otimes\boldsymbol{I}_d\right)\right\|$$
$$:=J_{T,61}+J_{T,62}+J_{T,63}+J_{T,64}.$$

We can show that $J_{T,62}$ and $J_{T,63}$ are $O_P(d_T)$, since

$$\max_{t\geq 1}\left\|\sum_{r=1}^{\infty}\widetilde{\mathbb{F}}_{r,t}(1)\right\| \leq \max_{t\geq 1}\sum_{r=1}^{\infty}\sum_{j=1}^{\infty}\left\|\widetilde{\mathbb{F}}_{rj,t}\right\| \leq \max_{t\geq 1}\sum_{r=1}^{\infty}\sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}\left\|\mathbb{B}_t^{r+k}(1)\right\|\left\|\mathbb{B}_t^{k}(1)\right\|$$
$$\leq \max_{t\geq 1}\left(\sum_{r=1}^{\infty}\left\|\mathbb{B}_t^{r}(1)\right\|\right)\left(\sum_{j=1}^{\infty}j\left\|\mathbb{B}_t^{j}(1)\right\|\right) \leq M\max_{t\geq 1}\sum_{j=1}^{\infty}j\sum_{k=0}^{\infty}\left\|\boldsymbol{B}_{k+j,t}\right\|\left\|\boldsymbol{B}_{k,t}\right\|$$
$$\leq M\max_{t\geq 1}\left(\sum_{j=1}^{\infty}j\left\|\boldsymbol{B}_{j,t}\right\|\right)\left(\sum_{k=0}^{\infty}\left\|\boldsymbol{B}_{k,t}\right\|\right) < \infty.$$

Similar to $J_{T,54}$, we have $J_{T,64} = o_P(1)$, since

$$\sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\left\|\widetilde{\mathbb{F}}_{r,t+1}(1)-\widetilde{\mathbb{F}}_{r,t}(1)\right\| \leq \sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}\left\|\mathbb{B}_{t+1}^{r+k}(1)\otimes\mathbb{B}_{t+1}^{k}(1)-\mathbb{B}_{t}^{r+k}(1)\otimes\mathbb{B}_{t}^{k}(1)\right\|$$
$$\leq \sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}\left(\left\|\mathbb{B}_{t+1}^{r+k}(1)-\mathbb{B}_{t}^{r+k}(1)\right\|\left\|\mathbb{B}_{t+1}^{k}(1)\right\|+\left\|\mathbb{B}_{t+1}^{k}(1)-\mathbb{B}_{t}^{k}(1)\right\|\left\|\mathbb{B}_{t}^{r+k}(1)\right\|\right)$$
$$\leq \left(\sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\left\|\mathbb{B}_{t+1}^{r}(1)-\mathbb{B}_{t}^{r}(1)\right\|\right)\left(\max_{t\geq 1}\sum_{j=1}^{\infty}j\left\|\mathbb{B}_t^{j}\right\|\right)$$
$$+\left(\sum_{t=1}^{T-1}\sum_{j=1}^{\infty}j\left\|\mathbb{B}_{t+1}^{j}(1)-\mathbb{B}_{t}^{j}(1)\right\|\right)\left(\max_{t\geq 1}\sum_{r=1}^{\infty}\left\|\mathbb{B}_t^{r}\right\|\right) = O(1).$$

Now consider term $J_{T,61}$. Define $\boldsymbol{u}_t = \sum_{r=1}^{\infty}\mathbb{F}_{r,t}(1)\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^{\top}\otimes\boldsymbol{I}_d\right)$, $\boldsymbol{u}_t' = \boldsymbol{u}_t I(\|\boldsymbol{u}_t\| \leq T^{\frac{2}{\delta}})$ and $\boldsymbol{u}_t'' = \boldsymbol{u}_t - \boldsymbol{u}_t'$. Then we have

$$\begin{aligned}J_{T,61} =& \max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}(s_l)\|\left(\boldsymbol{u}_t'+\boldsymbol{u}_t''-E(\boldsymbol{u}_t'+\boldsymbol{u}_t''|\mathcal{F}_{t-1})\right)\right\|\\ \leq& \max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}(s_l)\|\left(\boldsymbol{u}_t'-E(\boldsymbol{u}_t'|\mathcal{F}_{t-1})\right)\right\|+\max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}(s_l)\|\,\boldsymbol{u}_t''\right\|\\ &+\max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}(s_l)\|\,E(\boldsymbol{u}_t''|\mathcal{F}_{t-1})\right\| := J_{T,611}+J_{T,612}+J_{T,613}.\end{aligned}$$

Using the same argument as that used in the proof of $J_{T,12}$ in Lemma A.4, we can show that $J_{T,612}$ and $J_{T,613}$ are $O_P\left(T^{\frac{2}{\delta}}d_T\right)$. Next, consider $J_{T,611}$. For any $1\leq l\leq N_T$, let $\boldsymbol{Y}_t = \|\boldsymbol{W}_{T,t}(s_l)\|\left(\boldsymbol{u}_t'-E(\boldsymbol{u}_t'|\mathcal{F}_{t-1})\right)$. We then have $E\left(\boldsymbol{Y}_t|\mathcal{F}_{t-1}\right)=0$ and $\|\boldsymbol{Y}_t\|\leq 2T^{2/\delta}d_T$. In addition, we have

$$\max_{1\leq l\leq N_T}\left\|\sum_{t=1}^{T}E(\boldsymbol{Y}_t\boldsymbol{Y}_t^{\top}|\mathcal{F}_{t-1})\right\| \leq 4\max_{1\leq l\leq N_T}\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}(s_l)\|^2 E\left(\|\boldsymbol{u}_t\|^2|\mathcal{F}_{t-1}\right)$$
$$\leq M\cdot d_T\max_{1\leq l\leq N_T}\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}(s_l)\|\sum_{r=1}^{\infty}\|\mathbb{F}_{r,t}(1)\|^2\|\boldsymbol{\epsilon}_{t-r}\|^2 \leq M\cdot d_T\max_{t\geq 1}\sum_{r=1}^{\infty}\|\mathbb{F}_{r,t}(1)\|^2\|\boldsymbol{\epsilon}_{t-r}\|^2$$
$$\leq M\cdot d_T\sum_{r=1}^{\infty}\max_{t\geq 1}\|\mathbb{F}_{r,t}(1)\|^2\left(\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_{t-r}\|^{\delta}\right)^{\frac{2}{\delta}} = O_P\left(d_T T^{\frac{2}{\delta}}\right).$$

Therefore, we have $\max_{1\le l\le N_T}\left\|\sum_{t=1}^T E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| = O_P(d_T T^{\frac{2}{\delta}})$. By Lemma A.2, and choosing $\beta=4$, we have

$$
\begin{aligned}
&\Pr\left(J_{T,611} > \sqrt{\beta M}\sqrt{d_T T^{\frac{2}{\delta}}\log T}\right)\\
=&\Pr\left(J_{T,611} > \sqrt{\beta M}\sqrt{d_T T^{\frac{2}{\delta}}\log T}, \max_{1\le l\le N_T}\left\|\sum_{t=1}^T E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| \le M d_T T^{\frac{2}{\delta}}\right)\\
&+\Pr\left(J_{T,611} > \sqrt{\beta M}\sqrt{d_T T^{\frac{2}{\delta}}\log T}, \max_{1\le l\le N_T}\left\|\sum_{t=1}^T E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| > M d_T T^{\frac{2}{\delta}}\right)\\
\le&\Pr\left(J_{T,611} > \sqrt{\beta M}\sqrt{d_T T^{\frac{2}{\delta}}\log T}, \max_{1\le l\le N_T}\left\|\sum_{t=1}^T E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| \le M d_T T^{\frac{2}{\delta}}\right)\\
&+\Pr\left(\max_{1\le l\le N_T}\left\|\sum_{t=1}^T E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| > M d_T T^{\frac{2}{\delta}}\right)\\
\le& N_T\exp\left(-\frac{\beta M d_T T^{\frac{2}{\delta}}\log T}{2(M d_T T^{\frac{2}{\delta}} + M\sqrt{d_T T^{\frac{2}{\delta}}\log T T^{\frac{2}{\delta}} d_T})}\right) + o(1) \le N_T \exp\left(-\frac{\beta}{2}\log T\right) = N_T T^{-\frac{\beta}{2}} = o(1)
\end{aligned}
$$

given $d_T T^{\frac{4}{\delta}}\log T \to 0$. Hence, we have $J_{T,611} = O_P(\{d_T T^{\frac{2}{\delta}}\log T\}^{1/2})$. Combining the above results, we have proved that $\sup_{\tau\in[a,b]}\left|\sum_{t=1}^T \|\boldsymbol{W}_{T,t}(\tau)\| E\left(\|\boldsymbol{\zeta}_t\boldsymbol{\epsilon}_t\|^2|\mathcal{F}_{t-1}\right)\right| = O_P(1)$.

Finally, we turn to $J_{T,3}$, and apply the truncation method. Let $\boldsymbol{u}_t = \boldsymbol{\zeta}_t\boldsymbol{\epsilon}_t$, $\boldsymbol{u}_t' = \boldsymbol{u}_t I\left(\|\boldsymbol{u}_t\| \le T^{\frac{2}{\delta}}\right)$ and $\boldsymbol{u}_t'' = \boldsymbol{u}_t - \boldsymbol{u}_t'$. Then we have

$$
\begin{aligned}
J_{T,3} =& \max_{1\le l\le N_T}\left\|\sum_{t=1}^T (\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(s_l))\left(\boldsymbol{u}_t' + \boldsymbol{u}_t'' - E(\boldsymbol{u}_t'+\boldsymbol{u}_t''|\mathcal{F}_{t-1})\right)\right\|\\
\le& \max_{1\le l\le N_T}\left\|\sum_{t=1}^T (\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(s_l))\left(\boldsymbol{u}_t' - E(\boldsymbol{u}_t'|\mathcal{F}_{t-1})\right)\right\| + \max_{1\le l\le N_T}\left\|\sum_{t=1}^T (\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(s_l))\,\boldsymbol{u}_t''\right\|\\
&+\max_{1\le l\le N_T}\left\|\sum_{t=1}^T (\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(s_l))\,E(\boldsymbol{u}_t''|\mathcal{F}_{t-1})\right\| = J_{T,31}+J_{T,32}+J_{T,33}.
\end{aligned}
$$

It's easy to show that $J_{T,32} = O_P(T^{\frac{2}{\delta}}d_T)$ and $J_{T,33} = O_P(T^{\frac{2}{\delta}}d_T)$. Thus, we focus on $J_{T,31}$.

For any $1\le l\le N_T$, let $\boldsymbol{Y}_t = (\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(s_l))(\boldsymbol{u}_t' - E(\boldsymbol{u}_t'|\mathcal{F}_{t-1}))$, then we have $E\left(\boldsymbol{Y}_t|\mathcal{F}_{t-1}\right) = 0$ and $\|\boldsymbol{Y}_t\| \le 2T^{2/\delta}d_T$. Also,

$$
\max_{1\le l\le N_T}\sum_{t=1}^T \|\boldsymbol{W}_{T,t}(s_l)\| E\left(\|\boldsymbol{\zeta}_t\boldsymbol{\epsilon}_t\|^2|\mathcal{F}_{t-1}\right) = O_P(1),
$$

which yields

$$
\max_{1\le l\le N_T}\left\|\sum_{t=1}^T E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| \le M d_T \max_{1\le l\le N_T}\sum_{t=1}^T \|\boldsymbol{W}_{T,t}(s_l)\| E\left(\|\boldsymbol{u}_t\|^2|\mathcal{F}_{t-1}\right) = O_P\left(d_T\right).
$$

Therefore, we have $\max_{1\le l\le N_T}\left\|\sum_{t=1}^T E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| = O_P(d_T)$. By Lemma A.2 and choosing $\beta=4$, we have

$$
\begin{aligned}
&\Pr\left(J_{T,31} > \sqrt{\beta M}\gamma_T\right) = \Pr\left(J_{T,31} > \sqrt{\beta M}\gamma_T, \max_{1\le l\le N_T}\left\|\sum_{t=1}^{T} E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| \le Md_T\right)\\
&+\Pr\left(J_{T,31} > \sqrt{\beta M}\gamma_T, \max_{1\le l\le N_T}\left\|\sum_{t=1}^{T} E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| > Md_T\right)\\
&\le \Pr\left(J_{T,31} > \sqrt{\beta M}\gamma_T, \max_{1\le l\le N_T}\left\|\sum_{t=1}^{T} E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| \le Md_T\right)\\
&+\Pr\left(\max_{1\le l\le N_T}\left\|\sum_{t=1}^{T} E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| > Md_T\right) \le N_T\exp\left(-\frac{\beta M\gamma_T^2}{2(Md_T + M\gamma_T T^{\frac{2}{\delta}}d_T)}\right) + o(1)\\
&\le N_T\exp\left(-\frac{\beta}{2}\log T\right) = N_T T^{-\frac{\beta}{2}} = o(1)
\end{aligned}
$$

given $d_T T^{\frac{4}{\delta}}\log T \to 0$.

We now have completed the proof of the second result. □

Proof of Lemma A.6.

Let $\boldsymbol{\Psi}_j(\tau) = \boldsymbol{J}\boldsymbol{\Phi}^j(\tau)\boldsymbol{J}^\top$, where

$$
\boldsymbol{\Phi}(\tau) = \begin{pmatrix} \boldsymbol{A}_1(\tau) & \cdots & \boldsymbol{A}_{p-1}(\tau) & \boldsymbol{A}_p(\tau) \\ \boldsymbol{I}_d & \cdots & \boldsymbol{0}_d & \boldsymbol{0}_d \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0}_d & \cdots & \boldsymbol{I}_d & \boldsymbol{0}_d \end{pmatrix}
$$

and $\boldsymbol{J} = \left[\boldsymbol{I}_d, \boldsymbol{0}_{d\times d(p-1)}\right]$.

To proceed, we write $\boldsymbol{y_t}$ as a time–varying VMA(∞):

$$
\boldsymbol{y_t} = \sum_{j=0}^{\infty}\boldsymbol{\Psi}_{j,t}\left(\sum_{l=0}^{q}\boldsymbol{B}_l(\tau_{t-j})\boldsymbol{x}_{t-l-j} + \boldsymbol{\eta}_{t-j}\right) = \boldsymbol{\mu}_t + \sum_{j=0}^{\infty}\boldsymbol{D}_{j,t}^{\boldsymbol{\epsilon}}\boldsymbol{\epsilon}_{t-j} + \sum_{l=0}^{q}\sum_{j=0}^{\infty}\boldsymbol{D}_{j,l,t}^{\boldsymbol{v}}\boldsymbol{v}_{t-l-j},
$$

where $\boldsymbol{\mu}_t = \sum_{j=0}^{\infty}\sum_{l=0}^{q}\boldsymbol{\Psi}_{j,t}\boldsymbol{B}_l(\tau_{t-j})\boldsymbol{g}(\tau_{t-l-j})$, $\boldsymbol{\Psi}_{j,t} = \boldsymbol{J}\prod_{m=0}^{j-1}\boldsymbol{\Phi}(\tau_{t-m})\boldsymbol{J}^\top$, $\boldsymbol{D}_{j,t}^{\boldsymbol{\epsilon}} = \boldsymbol{\Psi}_{j,t}\boldsymbol{\omega}(\tau_{t-j})$, and $\boldsymbol{D}_{j,l,t}^{\boldsymbol{v}} = \sum_{k=0}^{j}\boldsymbol{\Psi}_{k,t}\boldsymbol{B}_l(\tau_{t-k})\boldsymbol{C}_{j-k}(\tau_{t-l-k})$.

Let ρ denote the largest eigenvalue of $\boldsymbol{\Phi}(\tau)$ uniformly over $\tau \in [0,1]$. Then we have $\rho < 1$ by Assumption 1.1. In addition, similar to the proof of Proposition 2.4 in Dahlhaus and Polonik (2009), we have $\max_{t\ge 1}\|\prod_{m=0}^{j-1}\boldsymbol{\Phi}(\tau_{t-m})\| \le M\rho^j$.

Next, we will show that $\boldsymbol{y_t}$ can be approximated by a time-varying MA(∞) process $\widetilde{\boldsymbol{y}}_t$ satisfying $\{E\|\boldsymbol{y_t} - \widetilde{\boldsymbol{y}}_t\|^\delta\}^{1/\delta} = O(T^{-1})$, where $\widetilde{\boldsymbol{y}}_t$ has been defined in the body of this lemma. It follows that

$$
\begin{aligned}
\{E\|\boldsymbol{y_t} - \widetilde{\boldsymbol{y}}_t\|^\delta\}^{1/\delta} &\le M\left(\|\boldsymbol{\mu}_t - \boldsymbol{\mu}(\tau_t)\| + \sum_{j=0}^{\infty}\left\|\boldsymbol{D}_{j,t}^{\boldsymbol{\epsilon}} - \boldsymbol{D}_{j}^{\boldsymbol{\epsilon}}(\tau_t)\right\| + \sum_{l=0}^{q}\sum_{j=0}^{\infty}\left\|\boldsymbol{D}_{j,l,t}^{\boldsymbol{v}} - \boldsymbol{D}_{j,l}^{\boldsymbol{v}}(\tau_t)\right\|\right)\\
&:= O(1)\cdot(I_{T,1} + I_{T,2} + I_{T,3}),
\end{aligned}
$$

where the definitions of $I_{T,1}$, $I_{T,2}$, and $I_{T,3}$ are obvious.

Consider $I_{T,1}$. Note that for any conformable matrices $\{\boldsymbol{A}_i\}$ and $\{\boldsymbol{B}_i\}$, since

$$\prod_{i=1}^{r}\boldsymbol{A}_i-\prod_{i=1}^{r}\boldsymbol{B}_i=\sum_{j=1}^{r}\left(\prod_{k=1}^{j-1}\boldsymbol{A}_k\right)(\boldsymbol{A}_j-\boldsymbol{B}_j)\left(\prod_{k=j+1}^{r}\boldsymbol{B}_k\right),$$

we obtain

$$\begin{aligned}\|\boldsymbol{\Psi}_{j,t}-\boldsymbol{\Psi}_j(\tau_t)\| &= \left\|\boldsymbol{J}\prod_{m=0}^{j-1}\boldsymbol{\Phi}(\tau_{t-m})\boldsymbol{J}^\top-\boldsymbol{J}\boldsymbol{\Phi}^j(\tau_t)\boldsymbol{J}^\top\right\| \\ &\leq M\sum_{i=1}^{j-1}\left\|\boldsymbol{\Phi}^i(\tau_t)(\boldsymbol{\Phi}(\tau_{t-i})-\boldsymbol{\Phi}(\tau_t))\prod_{m=i+1}^{j-1}\boldsymbol{\Phi}(\tau_{t-m})\right\|\leq M\sum_{i=1}^{j-1}\frac{i}{T}\rho^{j-1}.\end{aligned}$$

Hence, we have

$$\begin{aligned}I_{T,1} &\leq \sum_{j=0}^{\infty}\|\boldsymbol{\Psi}_{j,t}-\boldsymbol{\Psi}_j(\tau_t)\|\cdot\sum_{l=0}^{q}\|\boldsymbol{B}_l(\tau_{t-j})\boldsymbol{g}(\tau_{t-l-j})\| \\ &\quad+\sum_{j=0}^{\infty}\|\boldsymbol{\Psi}_j(\tau_t)\|\cdot\sum_{l=0}^{q}\|\boldsymbol{B}_l(\tau_{t-j})\boldsymbol{g}(\tau_{t-l-j})-\boldsymbol{B}_l(\tau_t)\boldsymbol{g}(\tau_t)\| \\ &\leq M\sum_{j=0}^{\infty}\sum_{i=1}^{j-1}\frac{i}{T}\rho^{j-1}+M\sum_{j=0}^{\infty}\rho^j\frac{j}{T}=O(T^{-1}),\end{aligned}$$

where we have used the facts that $\|\boldsymbol{\Psi}_j(\tau)\|\leq M\rho^j$ and

$$\begin{aligned}&\|\boldsymbol{B}_l(\tau_{t-j})\boldsymbol{g}(\tau_{t-l-j})-\boldsymbol{B}_l(\tau_t)\boldsymbol{g}(\tau_t)\| \\ &=\|\boldsymbol{B}_l(\tau_{t-j})\boldsymbol{g}(\tau_{t-l-j})-\boldsymbol{B}_l(\tau_t)\boldsymbol{g}(\tau_{t-l-j})+\boldsymbol{B}_l(\tau_t)\boldsymbol{g}(\tau_{t-l-j})-\boldsymbol{B}_l(\tau_t)\boldsymbol{g}(\tau_t)\| \\ &\leq\|\boldsymbol{B}_l(\tau_{t-j})-\boldsymbol{B}_l(\tau_t)\|\cdot\|\boldsymbol{g}(\tau_{t-l-j})\|+\|\boldsymbol{B}_l(\tau_t)\|\cdot\|\boldsymbol{g}(\tau_{t-l-j})-\boldsymbol{g}(\tau_t)\|\leq M\frac{j}{T}.\end{aligned}$$

Similarly, we have $I_{T,2}=O(T^{-1})$.

For $I_{T,3}$,

$$\begin{aligned}I_{T,3} &\leq \sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{j}\|\boldsymbol{\Psi}_{k,t}\boldsymbol{B}_l(\tau_{t-k})\boldsymbol{C}_{j-k}(\tau_{t-l-k})-\boldsymbol{\Psi}_k(\tau_t)\boldsymbol{B}_l(\tau_t)\boldsymbol{C}_{j-k}(\tau_t)\| \\ &=\sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\|\boldsymbol{\Psi}_{j,t}\boldsymbol{B}_l(\tau_{t-j})\boldsymbol{C}_k(\tau_{t-l-j})-\boldsymbol{\Psi}_j(\tau_t)\boldsymbol{B}_l(\tau_t)\boldsymbol{C}_k(\tau_t)\| \\ &\leq\sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\|\boldsymbol{\Psi}_{j,t}-\boldsymbol{\Psi}_j(\tau_t)\|\cdot\|\boldsymbol{B}_l(\tau_{t-j})\|\cdot\|\boldsymbol{C}_k(\tau_{t-l-j})\| \\ &\quad+\sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\|\boldsymbol{\Psi}_j(\tau_t)\|\cdot\|\boldsymbol{B}_l(\tau_{t-j})-\boldsymbol{B}_l(\tau_t)\|\cdot\|\boldsymbol{C}_k(\tau_{t-l-j})\| \\ &\quad+\sum_{l=0}^{q}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\|\boldsymbol{\Psi}_j(\tau_t)\|\cdot\|\boldsymbol{B}_l(\tau_t)\|\cdot\|\boldsymbol{C}_k(\tau_{t-l-j})-\boldsymbol{C}_k(\tau_t)\| \\ &\leq M\sum_{j=0}^{\infty}\sum_{i=1}^{j-1}\frac{i}{T}\rho^{j-1}\cdot\sup_{\tau\in[0,1]}\sum_{k=0}^{\infty}\|\boldsymbol{C}_k(\tau)\|\end{aligned}$$

$$+\frac{M}{T}\left(\sum_{j=0}^{\infty} j\rho^j\right)\left(\sup_{\tau\in[0,1]}\sum_{k=0}^{\infty}\|\boldsymbol{C}_k(\tau)\| + \sup_{\tau\in[0,1]}\sum_{k=0}^{\infty}\left\|\boldsymbol{C}_k^{(1)}(\tau)\right\|\right) = O(T^{-1}).$$

In addition, it is straightforward to verify that

$$\sup_{\tau\in[0,1]}\sum_{j=0}^{\infty} j\|\boldsymbol{D}_j^{\boldsymbol{\epsilon},(k)}(\tau)\| < \infty \quad \text{and} \quad \sup_{\tau\in[0,1]}\sum_{j=0}^{\infty} j\|\boldsymbol{D}_{j,l}^{\boldsymbol{v},(k)}(\tau)\| < \infty$$

for $k = 0, 1$ (see, the proof of Propositions 2.1, for example.)

The proof is now completed. □

Proof of Lemma A.7.

(1)-(2). To prove parts (1) and (2), it suffices to show that

$$\sup_{\tau\in[0,1]}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{Z}_t\boldsymbol{Z}_t^\top - E(\boldsymbol{Z}_t\boldsymbol{Z}_t^\top)\right)\left(\frac{\tau_t-\tau}{h}\right)^k K_h(\tau_t-\tau)\right\| = O_P\left(\sqrt{\frac{\log T}{Th}}\right)$$

for $k = 0, 1, 2$. Since $\boldsymbol{Z}_t$ can be approximated by a time-varying VMA(∞) process by Lemma A.6, then the uniform convergence results follows directly from Lemma 2.3.

(3). Part (3) follows directly from Lemma B.8 (1) in Yan et al. (2020).

□

Proof of Lemma A.8.

(1). By Lemma A.7,

$$\boldsymbol{S}\boldsymbol{X}_{\boldsymbol{C}} = \begin{pmatrix} (\boldsymbol{X}_{\widetilde{\boldsymbol{C}},1}^\top, \boldsymbol{0}_{d\times(l-s)})(\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau_1}^\top\boldsymbol{K}_{\tau_1}\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau_1})^{-1}\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau_1}^\top\boldsymbol{K}_{\tau_1}\boldsymbol{X}_{\boldsymbol{C}} \\ \vdots \\ (\boldsymbol{X}_{\widetilde{\boldsymbol{C}},T}^\top, \boldsymbol{0}_{d\times(l-s)})(\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau_T}^\top\boldsymbol{K}_{\tau_T}\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau_T})^{-1}\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau_T}^\top\boldsymbol{K}_{\tau_T}\boldsymbol{X}_{\boldsymbol{C}} \end{pmatrix}$$
$$= \begin{pmatrix} \boldsymbol{X}_{\widetilde{\boldsymbol{C}},1}^\top\boldsymbol{\Sigma}_{\boldsymbol{X}_{\widetilde{\boldsymbol{C}}}}^{-1}(\tau_1)\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}^\top(\tau_1) \\ \vdots \\ \boldsymbol{X}_{\widetilde{\boldsymbol{C}},T}^\top\boldsymbol{\Sigma}_{\boldsymbol{X}_{\widetilde{\boldsymbol{C}}}}^{-1}(\tau_T)\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}^\top(\tau_T) \end{pmatrix}(1+o_P(1)),$$

which follows that

$$\frac{1}{T}\boldsymbol{X}_{\boldsymbol{C}}^\top(\boldsymbol{I}_{dT}-\boldsymbol{S})^\top(\boldsymbol{I}_{dT}-\boldsymbol{S})\boldsymbol{X}_{\boldsymbol{C}}$$
$$= \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{X}_{\boldsymbol{C},t}^\top - \boldsymbol{X}_{\widetilde{\boldsymbol{C}},t}^\top\boldsymbol{\Sigma}_{\boldsymbol{X}_{\widetilde{\boldsymbol{C}}}}^{-1}(\tau_t)\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}^\top(\tau_t)\right)^\top\left(\boldsymbol{X}_{\boldsymbol{C},t}^\top - \boldsymbol{X}_{\widetilde{\boldsymbol{C}},t}^\top\boldsymbol{\Sigma}_{\boldsymbol{X}_{\widetilde{\boldsymbol{C}}}}^{-1}(\tau_t)\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}^\top(\tau_t)\right) + o_P(1).$$

Note that each element of $\boldsymbol{\Sigma}_{\boldsymbol{X}_{\widetilde{\boldsymbol{C}}}}(\tau)$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}(\tau)$ is Lipschitz continuous. Thus, by Lemmas 2.2 and A.6, the result holds.

(2). Let $\rho_T = h^2 + \sqrt{\log T/(Th)}$. By Lemma A.7.1, we have

$$[\boldsymbol{X}_{\widetilde{\boldsymbol{C}},t}^\top, \boldsymbol{0}_{d\times(l-s)}](\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau_t}^\top\boldsymbol{K}_{\tau_t}\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau_t})^{-1}\boldsymbol{X}_{\widetilde{\boldsymbol{C}},\tau_t}^\top\boldsymbol{K}_{\tau_t}\widetilde{\boldsymbol{X}} = \boldsymbol{X}_{\widetilde{\boldsymbol{C}},t}^\top\boldsymbol{\theta}(\tau_t)(1+O_P(\rho_T))$$

uniformly over $1 \leq t \leq T$. Hence, we have

$$\begin{aligned}
&\boldsymbol{X}_{\boldsymbol{C}}^{\top}(\boldsymbol{I}_{dT} - \boldsymbol{S})^{\top}(\boldsymbol{I}_{dT} - \boldsymbol{S})\widetilde{\boldsymbol{X}} \\
&= \sum_{t=1}^{T} \Big(\boldsymbol{X}_{\boldsymbol{C},t}\boldsymbol{X}_{\widetilde{\boldsymbol{C}},t}^{\top} - \boldsymbol{\Sigma}_{\boldsymbol{X}_{\boldsymbol{C},\widetilde{\boldsymbol{C}}}}(\tau_t)\boldsymbol{\Sigma}_{\boldsymbol{X}_{\widetilde{\boldsymbol{C}}}}^{-1}(\tau_t)\boldsymbol{X}_{\widetilde{\boldsymbol{C}},t}\boldsymbol{X}_{\widetilde{\boldsymbol{C}},t}^{\top}(1 + O_P(\rho_T)) \Big) \boldsymbol{\theta}(\tau_t) \cdot O_P(\rho_T) = O_P(T\rho_T^2),
\end{aligned}$$

where the last equality follows from Lemma 2.2. Finally, the result holds since $O_P(T\rho_T^2) = o_P(\sqrt{T})$ by Assumption 6. □

References

Benati, L. and Surico, P. (2009), 'VAR analysis and the great moderation', *American Economic Review* **99**(4), 1636–52.

Beveridge, S. and Nelson, C. R. (1981), 'A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle', *Journal of Monetary Economics* **7**(2), 151–174.

Bühlmann, P. (1998), 'Sieve bootstrap for smoothing in nonstationary time series', *Annals of Statistics* **26**(1), 48–83.

Chu, C.-K. and Marron, J. S. (1991), 'Comparison of two bandwidth selectors with dependent errors', *Annals of Statistics* **19**(4), 1906–1918.

Cogley, T., Primiceri, G. E. and Sargent, T. J. (2010), 'Inflation-gap persistence in the U.S.', *American Economic Journal: Macroeconomics* **2**(1), 43–69.

Dahlhaus, R. (1996), 'On the Kullback-Leibler information divergence of locally stationary processes', *Stochastic Processes and Their Applications* **62**(1), 139–168.

Dahlhaus, R. (2012), Locally stationary processes, *in* 'Handbook of Statistics', Vol. 30, Elsevier, pp. 351–413.

Dahlhaus, R. and Polonik, W. (2009), 'Empirical spectral processes for locally stationary time series', *Bernoulli* **15**(1), 1–39.

Dahlhaus, R. and Rao, S. S. (2006), 'Statistical inference for time-varying ARCH processes', *Annals of Statistics* **34**(3), 1075–1114.

Fan, J. and Huang, T. (2005), 'Profile likelihood inferences on semiparametric varying-coefficient partially linear models', *Bernoulli* **11**(6), 1031–1057.

Freedman, D. A. (1975), 'On tail probabilities for martingales', *Annals of Probability* **3**(1), 100–118.

Friedrich, M., Smeekes, S. and Urbain, J.-P. (2020), 'Autoregressive wild bootstrap inference for nonparametric trends', *Journal of Econometrics* **214**(1), 81–109.

Gao, J., Kanaya, S., Li, D. and Tjøstheim, D. (2015), 'Uniform consistency for nonparametric estimation in null recurrent time series', *Econometric Theory* **31**(4), 911–952.

Hall, P. G. and Racine, J. S. (2015), 'Infinite order cross-validated local polynomial regression', *Journal of Econometrics* **185**(2), 510–525.

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

Hamilton, J. D. (1994), *Time Series Analysis*, Vol. 2, Princeton University Press, New Jersey.

Hansen, B. E. (1991), 'Strong laws for dependent heterogeneous processes', *Econometric theory* **7**(2), 213–221.

Hansen, B. E. (1992), 'Consistent covariance matrix estimation for dependent heterogeneous processes', *Econometrica* pp. 967–972.

Hansen, B. E. (2001), 'The new econometrics of structural change: dating breaks in US labour productivity', *Journal of Economic Perspectives* **15**(4), 117–128.

Hansen, B. E. (2008), 'Uniform convergence rates for kernel estimation with dependent data', *Econometric Theory* **24**(3), 726–748.

Li, D., Phillips, P. C. B. and Gao, J. (2016), 'Uniform consistency of nonstationary kernel-weighted sample covariances for nonparametric regression', *Econometric Theory* **32**(3), 655–685.

Li, Q. and Racine, J. (2007), *Nonparametric Econometrics Theory and Practice*, Princeton University Press, New Jersey.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer Science & Business Media.

Newey, W. K. and West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(3), 703–708.

Palm, F. C., Smeekes, S. and Urbain, J.-P. (2011), 'Cross-sectional dependence robust block bootstrap panel unit root tests', *Journal of Econometrics* **163**(1), 85–104.

Paul, P. (2020), 'The time-varying effect of monetary policy on asset prices', *Review of Economics and Statistics* **102**(4), 690–704.

Petrova, K. (2019), 'A quasi-Bayesian local likelihood approach to time varying parameter VAR models', *Journal of Econometrics* **212**(1), 286–306.

Phillips, P. C. B. and Lee, J. H. (2013), 'Predictive regression under various degrees of persistence and robust long-horizon regression', *Journal of Econometrics* **177**(2), 250–264.

Phillips, P. C. B., Li, D. and Gao, J. (2017), 'Estimating smooth structural change in cointergration models', *Journal of Econometrics* **196**(1), 180–195.

Phillips, P. C. B. and Solo, V. (1992), 'Asymptotics for linear processes', *Annals of Statistics* **20**(2), 971–1001.

Plagborg-Møller, M. and Wolf, C. K. (2021), 'Local projections and vars estimate the same impulse responses', *Econometrica* **89**(2), 955–980.

Primiceri, G. E. (2005), 'Time varying structural vector autoregressions and monetary policy', *Review of Economic Studies* **72**(3), 821–852.

Primiceri, G. E. (2006), 'Why inflation rose and fell: policy-makers' beliefs and us postwar stabilization policy', *Quarterly Journal of Economics* **121**(3), 867–901.

Richter, S. and Dahlhaus, R. (2019), 'Cross validation for locally stationary processes', *Annals of Statistics* **47**(4), 2145–2173.

Shao, X. (2010), 'The dependent wild bootstrap', *Journal of the American Statistical Association* **105**(489), 218–235.

Stock, J. H. and Watson, M. W. (2016), 'Core inflation and trend inflation', *Review of Economics and Statistics* **98**(4), 770–784.

Xia, Y. (1998), 'Bias-corrected confidence bands in nonparametric regression', *Journal of the Royal Statistical Society: Series B* **60**(4), 797–811.

Yan, Y., Gao, J. and Peng, B. (2020), A class of time-varying vector moving average (∞) models. Working paper available at https://ideas.repec.org/p/msh/ebswps/2020-39.html.

Zhang, T. and Wu, W. B. (2011), 'Testing parametric assumptions of trends of a nonstationary time series', *Biometrika* **98**(3), 599–614.

Zhang, T. and Wu, W. B. (2012), 'Inference of time-varying regression models', *Annals of Statistics* **40**(3), 1376–1402.