Department of Econometrics and Business Statistics

A Class of Time-Varying Vector Moving Average (∞) Models

Yayi Yan, Jiti Gao and Bin Peng

October 2020

A Class of Time–Varying Vector Moving Average (∞) Models

Yayi Yan, Jiti Gao and Bin Peng

Monash University

September 15, 2020

Abstract

Multivariate time series analyses are widely encountered in practical studies, e.g., modelling policy transmission mechanism and measuring connectedness between economic agents. To better capture the dynamics, this paper proposes a class of multivariate dynamic models with time–varying coefficients, which have a general time–varying vector moving average (VMA) representation, and nest, for instance, time–varying vector autoregression (VAR), time–varying vector autoregression moving–average (VARMA), and so forth as special cases. The paper then develops a unified estimation method for the unknown quantities before an asymptotic theory for the proposed estimators is established. In the empirical study, we investigate the transmission mechanism of monetary policy using U.S. data, and uncover a fall in the volatilities of exogenous shocks. In addition, we find that (i) monetary policy shocks have less influence on inflation before and during the so–called Great Moderation, (ii) inflation is more anchored recently, and (iii) the long–run level of inflation is below, but quite close to the Federal Reserve's target of two percent after the beginning of the Great Moderation period.

Keywords: Multivariate Time Series Model; Nonparametric Kernel Estimation; Trending Stationarity

1   Introduction

Vector autoregressions (VARs), as well as their extensions like vector autoregressive moving average (VARMA) models and VARs with exogenous variables (VARX), are among some of the most popular frameworks for modelling dynamic interactions among multiple variables. These models arise mainly as a response to the "incredible" identification conditions embedded in the large–scale macroeconomic models (Sims, 1980). VAR modelling begins with minimal restrictions on the multivariate dynamic models. Gradually armed with identification information, VARs plus their statistical tool–kits like impulse response functions, are powerful approaches for conducting policy analysis. Also, VARs can be applied to other important tasks including data description and forecasting, see Stock and Watson (2001) for a detailed review. Despite the popularity, linear VAR models can always be rejected by data in empirical studies (cf., Tsay, 1998). For example, Stock and Watson (2016b) point out, "changes associated with the Great Moderation go beyond reduction in variances to include changes in dynamics and reduction in predictability."

To go beyond linear VAR models, various parametric time–varying VAR models have been proposed (e.g., Tsay, 1998, Sims and Zha, 2006, and references therein) in order to allow for abrupt structural breaks in economic relationships and obtain efficient estimation. However, model misspecification and parameter instability may undermine the performance of parametric time–varying VAR models. Usually, researchers do not know the true functional forms of the time–varying parameters, so the choice on the functional forms might be somewhat arbitrary. In addition, as pointed out by Hansen (2001), "it may seem unlikely that a structural break could be immediate and might seem more reasonable to allow a structural change to take a period of time to take effect". Therefore, it is reasonable to allow smooth structural changes over a period of time rather than abrupt structural changes. Another strand of the VAR literature assumes that structural coefficients evolve in a random way, such as Primiceri (2005) and Giannone et al. (2019). Recently, Giraitis et al. (2014) point out that "the theoretical asymptotic properties of estimating such processes via the Kalman, or related filters are unclear". Along this line, Giraitis et al. (2014) and Giraitis et al. (2018) have achieved some useful asymptotic results. However, estimation theory for time–varying impulse response functions, which are of interest in interpreting multivariate dynamic models, are not yet established in the random walk case.

It is worth pointing out that although nonparametric time–varying models have received much attention over the past three decades (Robinson, 1989; Dahlhaus and Rao, 2006; Cai, 2007; Zhang and Wu, 2012; Li et al., 2019; Richter et al., 2019, to name a few) for univariate time series, little study about multivariate dynamic models with deterministic time–varying coefficients has been done. One exception is Dahlhaus et al. (1999), who use a wavelet method to transform the time–varying VAR model to a linear approximation with an orthonormal wavelet basis, and show that the wavelet estimator attains the usual near–optimal minimax rate of L_2 convergence. However,

the asymptotic distributions and the order of dynamic lags are not studied therein. The paper thus aims to provide a set of inferential theories which can boost the applicability of the time–varying VAR models practically. Up to this point, it is worth bringing up the terminology "local stationarity", which dates back to the seminal work by Dahlhaus (1996) at least. Although several studies have been conducted along this line (Dahlhaus, 1996 and Zhang and Wu, 2012 on time–varying AR; Dahlhaus and Polonik, 2009 on time–varying ARMA; Rohan and Ramanathan, 2013 and Truquet, 2017 on time–varying ARCH/GARCH), the literature has not ventured much outside the univariate case. A commonly used method is to approximate a locally stationary process by a stationary approximation on each of the segments (Dahlhaus et al., 2019). However, it remains unclear to us how to extend this method to the multivariate case where the segments on which stationarity approximations for each univariate time series may be different.

One contribution of this paper is to show the versatility of an alternative approach that is especially designed for multivariate time–varying MA(∞) processes. Our approach relies on an explicit decomposition of time–varying MA(∞) processes into 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 in the decomposition yields a martingale approximation to the sum of time–varying MA(∞) processes. This paper then develops a unified asymptotic theory by utilizing some relevant properties for martingale sequences. We are then able to deal with a wide class of multivariate dynamic models with smooth time–varying coefficients, which have a general time–varying vector moving average (VMA) representation, nesting VAR, VARMA, VARX and so forth as special cases. Specifically, the structural coefficients are unknown functions of the re–scaled time, so that the proposed models can better capture the simultaneous relations among variables of interest over time. Such a modelling strategy is especially useful for analysing time series over a long horizon, since it offers a comprehensive treatment on tracking interests which are affected by frequently updated policies, environment, system, etc. Consider the empirical study of Section 4 as an example. In an economy system consisting of inflation, unemployment and interest rates, one priority is inferring time–varying impacts of the interest rate change, which helps stabilize fluctuations in inflation and unemployment in long–run. Under the proposed framework, it is achieved by investigating the corresponding time–varying impulse response functions.

In summary, our contributions are in threefold. First, we propose a new class of time–varying VMA(∞) models, which covers several classes of multivariate dynamic models. Second, we establish the corresponding asymptotic theory for the proposed models and estimators by utilizing a time–varying counterpart of the conventional BN decomposition. Third, in the empirical study of Section 4, we study the changing dynamics of three key U.S. macroeconomic variables (i.e., inflation, unemployment, and interest rate), and uncover a fall in the volatilities of exogenous shocks. In addition, we find that (i) monetary policy shocks have less influence on inflation before and during the so–called Great Moderation; (ii) inflation is more anchored recently; and (iii) the long–run

2

level of inflation is below, but quite close to the Federal Reserve's target of two percent after the beginning of the Great Moderation period.

The organization of this paper is as follows. Section 2 proposes a class of time–varying VMA(∞) models. In Section 3, we discuss a class of time–varying VAR models and then establish an estimation theory for the unknown quantities. Section 4 presents an empirical study to show the practical relevance and applicability of the proposed models and estimation theory. Section 5 gives a short conclusion. The main proofs of the theorems are given in Appendix A. In the online supplementary material, simulation results are given in Appendix B.1 and some technical lemmas and proofs are given in the rest of 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; $\boldsymbol{I}_a$ and $\boldsymbol{0}_a$ are $a \times a$ identity and null matrices respectively, and $\boldsymbol{0}_{a \times b}$ stands for a $a \times b$ matrix of zeros; for a function $g(w)$, let $g^{(j)}(w)$ be the jth 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 $\tilde{c}_k = \int_{-1}^{1} u^k K(u) du$ and $\tilde{v}_k = \int_{-1}^{1} u^k K^2(u) du$ for integer $k \geq 0$; vec($\cdot$) stacks the elements of an $m \times n$ matrix as an $mn \times 1$ vector; for any $a \times a$ square matrix $\boldsymbol{A}$, vech($\boldsymbol{A}$) denotes the $\frac{1}{2}a(a+1) \times 1$ vector obtained from vec($\boldsymbol{A}$) by eliminating all supra–diagonal elements of $\boldsymbol{A}$; 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 Structure of Time–Varying VMA(∞)

We start our investigation by considering a class of time–varying VMA(∞) model:

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

for $t = 1, \ldots, T$, where $\boldsymbol{x}_t$ is a d–dimensional vector of observable variables, $\boldsymbol{\mu}_t$ is a d–dimensional unknown trending function, $\boldsymbol{\epsilon}_t$ is a vector of d–dimensional random innovations, and d is fixed throughout this paper. Moreover, $\mathbb{B}_t(L) = \sum_{j=0}^{\infty} \boldsymbol{B}_{j,t} L^j$, where L is the lag operator, and $\boldsymbol{B}_{j,t}$ is a matrix of $d \times d$ unknown deterministic coefficients.

We first comment on the usefulness of the structure in (2.1), and the corresponding BN decomposition. An application of the BN decomposition gives

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

where we have used the decomposition of $\mathbb{B}_t(L)$ as $\mathbb{B}_t(L) = \mathbb{B}_t(1) - (1-L)\widetilde{\mathbb{B}}_t(L)$, in which $\widetilde{\mathbb{B}}_t(L) = \sum_{j=0}^{\infty} \widetilde{\boldsymbol{B}}_{j,t} L^j$ and $\widetilde{\boldsymbol{B}}_{j,t} = \sum_{k=j+1}^{\infty} \boldsymbol{B}_{k,t}$. Equation (2.2) indicates that one may establish some general asymptotic properties for partial sums and quadratic forms of $\boldsymbol{x}_t$ with minor restrictions on $\{\boldsymbol{\epsilon}_t\}$.

For example, one can show that the simple average of $\boldsymbol{x}_t$ becomes

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_t = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\mu}_t + \frac{1}{T}\sum_{t=1}^{T}\mathbb{B}_t(1)\boldsymbol{\epsilon}_t + O_P(T^{-1}). \tag{2.3}$$

Similar to (2.3), asymptotic properties for partial sums and quadratic forms mainly depend on the probabilistic structure of $\{\boldsymbol{\epsilon}_t\}$ and regularity conditions on $\{\boldsymbol{B}_{j,t}\}$. In other words, there is no need to impose any further structure on $\boldsymbol{x}_t$, such as requiring $\boldsymbol{x}_t$ to be locally stationary time series in a similar way to what has been done in the relevant literature for the univariate case. As a consequence, it facilitates to develop general theory for the multivariate case. Moreover, it should be added that our settings in (2.2) and (2.3) considerably extend similar treatments by Phillips and Solo (1992) for the univariate linear process case where both $\boldsymbol{\mu}_t$ and $\boldsymbol{B}_{j,t}$ reduce to constant scalars: $\boldsymbol{\mu}_t = \mu$ and $\boldsymbol{B}_{j,t} = B_j$.

2.1 Examples and Useful Properties

Let us now stress that (2.1) covers a wide range of models, which are of general interest in both theory and practice. Below, we list a few examples, of which the parametric counterparts can be seen in Lütkepohl (2005).

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

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

which has been widely studied in the literature with Bayesian framework being the dominant approach (e.g., Benati and Surico, 2009; Paul, 2019). Similar to Hamilton (1994, p. 260), (2.4) can be expressed as a time–varying MA(∞) process $\boldsymbol{x}_t = \sum_{j=0}^{\infty}\boldsymbol{B}_{j,t}\boldsymbol{\epsilon}_{t-j}$, where $\boldsymbol{B}_{0,t} = \boldsymbol{I}_d$ and $\boldsymbol{B}_{j,t} = \boldsymbol{J}\prod_{i=0}^{j-1}\boldsymbol{\Phi}_{t-i}\boldsymbol{J}^{\top}$ for $j \geq 1$ with $\boldsymbol{J} = [\boldsymbol{I}_d, \boldsymbol{0}_{d\times d(p-1)}]$ and

$$\boldsymbol{\Phi}_t = \begin{pmatrix} \boldsymbol{A}_{1,t} & \cdots & \boldsymbol{A}_{p-1,t} & \boldsymbol{A}_{p,t} \\ \boldsymbol{I}_d & \cdots & \boldsymbol{0}_d & \boldsymbol{0}_d \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0}_d & \cdots & \boldsymbol{I}_d & \boldsymbol{0}_d \end{pmatrix}.$$

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

$$\boldsymbol{x}_t = \boldsymbol{A}_{1,t}\boldsymbol{x}_{t-1} + \dots + \boldsymbol{A}_{p,t}\boldsymbol{x}_{t-p} + \boldsymbol{\epsilon}_t + \boldsymbol{\Theta}_{1,t}\boldsymbol{\epsilon}_{t-1} + \dots + \boldsymbol{\Theta}_{q,t}\boldsymbol{\epsilon}_{t-q}, \tag{2.5}$$

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

Example 3. Let $\boldsymbol{x}_t$ be a d–dimensional time–varying double MA(∞) process:

$$\boldsymbol{x}_t = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_{j,t} \boldsymbol{v}_{t-j} \quad \text{and} \quad \boldsymbol{v}_t = \sum_{l=0}^{\infty} \boldsymbol{\Theta}_{l,t} \boldsymbol{\epsilon}_{t-l}, \tag{2.6}$$

in which the innovations $\boldsymbol{v}_t$'s also follow a time–varying MA(∞) process. Simple algebra shows that $\boldsymbol{x}_t = \sum_{j=0}^{\infty} \boldsymbol{B}_{j,t} \boldsymbol{\epsilon}_{t-j}$, where $\boldsymbol{B}_{j,t} = \sum_{l=0}^{j} \boldsymbol{\Psi}_{l,t} \boldsymbol{\Theta}_{j-l,t-l}$.

To facilitate the development of our general theory, we introduce the following assumptions.

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

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* $\delta > 2$.

Assumption 1 regulates the matrices $\boldsymbol{B}_{j,t}$'s, and ensures the validity of the BN decomposition under the time–varying framework. It covers cases such as (i) the parametric setting of Phillips and Solo (1992), and (ii) $\boldsymbol{B}_{j,t} := \boldsymbol{B}_j(\tau_t)$, where $\boldsymbol{B}_j(\cdot)$ satisfies Lipschitz continuity on $[0,1]$ for all j. 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 present a summary of useful results for Examples 1–3, which explains why model (2.1) serves as a foundation of the examples given above.

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,* $\limsup_{T\to\infty} \sum_{t=1}^{T-1} \|\boldsymbol{\Theta}_{m,t+1} - \boldsymbol{\Theta}_{m,t}\| < \infty$ *for* $m = 1, \ldots, q$. *Then both (2.5) and (2.6) are time–varying MA(∞) processes, in which the MA coefficients satisfy Assumption 1.*

2. *For Example 3, let* $\limsup_{T\to\infty} \sum_{t=1}^{T-1} \sum_{j=1}^{\infty} j\|\boldsymbol{\Psi}_{j,t+1} - \boldsymbol{\Psi}_{j,t}\| < \infty$ *and* $\max_{t\geq 1} \sum_{j=1}^{\infty} j\|\boldsymbol{\Psi}_{j,t}\| < \infty$. *Moreover, let* $\{\boldsymbol{\Theta}_{j,t}\}$ *satisfy the same conditions as those for* $\{\boldsymbol{\Psi}_{j,t}\}$. *Then (2.6) is a time–varying MA(∞), in which the MA coefficients satisfy Assumption 1.*

We now move on to investigate asymptotic properties for model (2.1). We first propose some estimates for several population moments of $\boldsymbol{x}_t$ in (2.1), which help derive the asymptotic theory throughout this paper. To conserve space, we present the rates of the uniform convergence below, while extra results on point–wise convergence are given in the supplementary Appendix B.

Theorem 2.1. *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 functions, in which m is fixed, 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 $d_T = \sup_{\tau \in [a,b], t \geq 1} \|\boldsymbol{W}_{T,t}(\tau)\|$.*

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. *$\sup_{\tau \in [a,b]} \|\sum_{t=1}^T \boldsymbol{W}_{T,t}(\tau)(\boldsymbol{x}_t \boldsymbol{x}_{t+p}^\top - E(\boldsymbol{x}_t \boldsymbol{x}_{t+p}^\top))\| = O_P(\sqrt{d_T \log T})$ 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..*

Theorem 2.1 is readily used for studying the most adopted case — the kernel weight (see Lemma B.7 of Appendix B for example), and will be repeatedly used in the theoretical derivation of the paper. Theorem 2.1 is also helpful to a broad range of studies such as those mentioned in Härdle et al. (2000), Fan and Yao (2003), Gao (2007), Li and Racine (2007), Hansen (2008), Wang and Chan (2014), and Li et al. (2016).

2.2 On the Trending Term — $\boldsymbol{\mu}_t$

As modelling time–varying means is always an important task in time series analysis (e.g., Wu and Zhao, 2007; Friedrich et al., 2020), we infer $\boldsymbol{\mu}_t$ of the model (2.1) below. Up to this point, we have not imposed any specific form on the components $\boldsymbol{\mu}_t$ and $\boldsymbol{B}_{j,t}$ of $\boldsymbol{x}_t$. To carry on with our investigation, we suppose further that $\boldsymbol{\mu}_t = \boldsymbol{\mu}(\tau_t)$ and $\boldsymbol{B}_{j,t} = \boldsymbol{B}_j(\tau_t)$ with $\tau_t = t/T$, so (2.1) can be expressed by

$$\boldsymbol{x}_t = \boldsymbol{\mu}(\tau_t) + \sum_{j=0}^\infty \boldsymbol{B}_j(\tau_t)\boldsymbol{\epsilon}_{t-j}. \tag{2.7}$$

The challenge then lies in the fact that "residuals" are time–varying linear processes. Some detailed explanations can be found in Dahlhaus (2012).

The following assumptions are necessary for the development of our trend estimation.

Assumption 3. *$\boldsymbol{\mu}(\cdot)$ and $\boldsymbol{B}_j(\cdot)$ are $d \times 1$ vector and $d \times d$ matrix respectively. Each functional component of $\boldsymbol{\mu}(\cdot)$ and $\boldsymbol{B}_j(\cdot)$ 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 probability kernel function and Lipschitz continuous on $[-1,1]$. Also, let $h \to 0$ and $Th \to \infty$ as $T \to \infty$.*

6

Assumption 3 can be considered as a stronger version than Assumption 1. Assumption 4 is standard in the literature of kernel regression (Li and Racine, 2007).

For $\forall \tau \in (0,1)$, we recover $\boldsymbol{\mu}(\tau)$ by the next estimator.

$$\widehat{\boldsymbol{\mu}}(\tau) = \left(\sum_{t=1}^{T} K_h(\tau_t - \tau)\right)^{-1} \sum_{t=1}^{T} \boldsymbol{x}_t K_h(\tau_t - \tau). \tag{2.8}$$

The corresponding asymptotic result is summarized in the next theorem.

Theorem 2.2. *Let Assumptions 2–4 hold. For $\forall \tau \in (0,1)$, as $T \to \infty$,*

$$\sqrt{Th}\left(\widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{\mu}(\tau) - \frac{1}{2}h^2\tilde{c}_2\boldsymbol{\mu}^{(2)}(\tau)\right) \to_D N\left(\mathbf{0}_{d\times 1}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau)\right),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau) = \tilde{v}_0\{\sum_{j=0}^{\infty}\boldsymbol{B}_j(\tau)\}\{\sum_{j=0}^{\infty}\boldsymbol{B}_j^{\top}(\tau)\}$, and $\tilde{c}_2$ and $\tilde{v}_0$ are defined in the end of Section 1.

Note that $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\cdot)$ is the long–run covariance matrix, which in general cannot be estimated directly. To construct confidence intervals practically, we use a dependent wild bootstrap (DWB) method which is initially proposed by Shao (2010) for stationary time series. For the sake of space, the detailed procedure with the associated asymptotic properties is presented in Appendix A.2.

3 Time–Varying VAR

In this section, we pay particular attention to one of the most popular models of the VMA(∞) family — VAR. Many multivariate time series exhibit time–varying simultaneous interrelationships and changes in unconditional volatility (e.g., Justiniano and Primiceri, 2008; Coibion and Gorodnichenko, 2011). Along this line, time–varying VAR models have proven to be especially useful for describing the dynamics of the multivariate time series. Majority time–varying VAR models are investigated under the Bayesian framework, while little has been done using a frequentists' approach. Building on Section 2, we consider a time–varying VAR model under the nonparametric framework, and establish the corresponding estimation theory.

Suppose that we observe $(\boldsymbol{x}_{-p+1}, \ldots, \boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ from the following data generating process. Accounting for heteroscedasticity, we consider the next model.

$$\boldsymbol{x}_t = \boldsymbol{a}(\tau_t) + \sum_{j=1}^{p} \boldsymbol{A}_j(\tau_t)\boldsymbol{x}_{t-j} + \boldsymbol{\eta}_t, \quad \text{with} \quad \boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau_t)\boldsymbol{\epsilon}_t, \tag{3.1}$$

where $\boldsymbol{\omega}(\tau)$ is a matrix of unknown functions of τ. The model (3.1) allows dynamic variations for both the coefficients and the covariance matrix. We infer $\boldsymbol{a}(\cdot)$ and $\boldsymbol{A}_j(\cdot)$'s below, which are respectively $d \times 1$ vector and $d \times d$ matrices of unknown smooth functions. In addition, we are interested in the $d \times d$ dimension $\boldsymbol{\omega}(\cdot)$ which governs the dynamics of the covariance matrix of

$\{\boldsymbol{\eta}_t\}_{t=1}^T$. As mentioned in Primiceri (2005), allowing $\boldsymbol{\omega}(\cdot)$ to vary over time is important theoretically and practically, because a constant covariance matrix implies that the shock to the ith variable of $\boldsymbol{x}_t$ has a time–invariant effect on the jth variable of $\boldsymbol{x}_t$, restricting simultaneous interactions among multiple variables to be time–invariant.

3.1 Estimation Method and Asymptotic Theory

To facilitate the development, we first impose the following conditions.

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 $\boldsymbol{A}(\tau) = (\boldsymbol{a}(\tau), \boldsymbol{A}_1(\tau), \ldots, \boldsymbol{A}_p(\tau))$ is second order continuously differentiable on $[0, 1]$ and $\boldsymbol{A}(\tau) = \boldsymbol{A}(0)$ for $\tau < 0$.*

3. *Each element 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 $\tau < 0$.*

Assumption 5.1 ensures that model (3.1) is neither unit–root nor explosive, while Assumption 5.2 allows the underlying data generate process to evolve over time in a smooth manner. In addition, the conditions $\boldsymbol{A}(\tau) = \boldsymbol{A}(0)$ and $\boldsymbol{\omega}(\tau) = \boldsymbol{\omega}(0)$ for $\tau < 0$ yield

$$\boldsymbol{x}_t = \boldsymbol{a}(0) + \sum_{j=1}^p \boldsymbol{A}_j(0)\boldsymbol{x}_{t-j} + \boldsymbol{\omega}(0)\boldsymbol{\epsilon}_t \tag{3.2}$$

for all $t \leq 0$, which ensures (3.2) behaves like a stationary parametric VAR(p) model for $t \leq 0$. Similar treatments can be found in Vogt (2012) for the univariate setting. With the above assumptions in hand, the following proposition shows that model (3.1) can be approximated by a time–varying VMA(∞) process satisfying Assumption 3.

Proposition 3.1. *Under Assumption 5, there exists a VMA(∞) process*

$$\widetilde{\boldsymbol{x}}_t = \boldsymbol{\mu}(\tau_t) + \boldsymbol{B}_0(\tau_t)\boldsymbol{\epsilon}_t + \boldsymbol{B}_1(\tau_t)\boldsymbol{\epsilon}_{t-1} + \boldsymbol{B}_2(\tau_t)\boldsymbol{\epsilon}_{t-2} + \cdots \tag{3.3}$$

such that $\max_{t \geq 1} E \|\boldsymbol{x}_t - \widetilde{\boldsymbol{x}}_t\| = O(T^{-1})$, where $\boldsymbol{\mu}(\tau) = \boldsymbol{a}(\tau) + \sum_{j=1}^\infty \boldsymbol{\Psi}_j(\tau)\boldsymbol{a}(\tau)$, $\boldsymbol{B}_0(\tau) = \boldsymbol{\omega}(\tau)$, $\boldsymbol{B}_j(\tau) = \boldsymbol{\Psi}_j(\tau)\boldsymbol{\omega}(\tau)$, $\boldsymbol{\Psi}_j(\tau) = \boldsymbol{J}\boldsymbol{\Phi}^j(\tau)\boldsymbol{J}^\top$ for $j \geq 1$, $\boldsymbol{\Phi}(\cdot)$ is defined as follows:

$$\boldsymbol{\Phi}(\tau) = \begin{pmatrix} \boldsymbol{A}_1(\tau) & \cdots & \boldsymbol{A}_{p-1}(\tau) & \boldsymbol{A}_p(\tau) \\ \boldsymbol{I}_d & \cdots & \boldsymbol{0}_d & \boldsymbol{0}_d \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0}_d & \cdots & \boldsymbol{I}_d & \boldsymbol{0}_d \end{pmatrix}, \tag{3.4}$$

and $\boldsymbol{J} = \left[\boldsymbol{I}_d, \boldsymbol{0}_{d \times d(p-1)} \right]$. *Moreover,* $\boldsymbol{\mu}(\cdot)$ *and* $\boldsymbol{B}_j(\cdot)$ *fulfil Assumption 3.*

Under Assumption 5, when τ_t is in a small neighbourhood of τ, we have

$$\boldsymbol{x}_t \approx \boldsymbol{Z}_{t-1}^\top \text{vec}[\boldsymbol{A}(\tau)] + \boldsymbol{\eta}_t, \tag{3.5}$$

where $\boldsymbol{Z}_{t-1} = \boldsymbol{z}_{t-1} \otimes \boldsymbol{I}_d$ and $\boldsymbol{z}_{t-1} = (1, \boldsymbol{x}_{t-1}^\top, \ldots, \boldsymbol{x}_{t-p}^\top)^\top$. The estimators of $\boldsymbol{A}(\tau)$ and $\boldsymbol{\Omega}(\tau)$ are then sequentially given by

$$\text{vec}[\widehat{\boldsymbol{A}}(\tau)] = \left(\sum_{t=1}^T \boldsymbol{Z}_{t-1} \boldsymbol{Z}_{t-1}^\top K_h(\tau_t - \tau) \right)^{-1} \sum_{t=1}^T \boldsymbol{Z}_{t-1} \boldsymbol{x}_t K_h(\tau_t - \tau),$$

$$\widehat{\boldsymbol{\Omega}}(\tau) = \left(\sum_{t=1}^T K_h(\tau_t - \tau) \right)^{-1} \sum_{t=1}^T \widehat{\boldsymbol{\eta}}_t \widehat{\boldsymbol{\eta}}_t^\top K_h(\tau_t - \tau), \tag{3.6}$$

where $\widehat{\boldsymbol{\eta}}_t = \boldsymbol{x}_t - \widehat{\boldsymbol{A}}(\tau_t) \boldsymbol{z}_{t-1}$. The asymptotic properties associated with (3.6) are summarized in the next theorem.

Theorem 3.1. *Let Assumptions 2, 4 and 5 hold. Suppose further that* $\frac{T^{1-\frac{4}{\delta}} h}{\log T} \to \infty$ *as* $T \to \infty$ *and* $\max_{t \geq 1} E(\|\boldsymbol{\epsilon}_t\|^4 | \mathcal{F}_{t-1}) < \infty$ *a.s. Then the following results hold.*

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

2. *In addition, conditional on* $\mathcal{F}_{t-1}$, *the third and fourth moments of* $\boldsymbol{\epsilon}_t$ *are identical to the corresponding unconditional moments a.s.. For* $\forall \tau \in (0, 1)$,

$$\sqrt{Th} \begin{pmatrix} \text{vec} \left(\widehat{\boldsymbol{A}}(\tau) - \boldsymbol{A}(\tau) - \frac{1}{2} h^2 \tilde{c}_2 \boldsymbol{A}^{(2)}(\tau) \right) \\ \text{vech} \left(\widehat{\boldsymbol{\Omega}}(\tau) - \boldsymbol{\Omega}(\tau) - \frac{1}{2} h^2 \tilde{c}_2 \boldsymbol{\Omega}^{(2)}(\tau) \right) \end{pmatrix} \to_D N \left(\boldsymbol{0}, \boldsymbol{V}(\tau) \right),$$

where $\boldsymbol{V}(\tau)$ *is defined in (A.3) for the sake of presentation.*

3. $\widehat{\boldsymbol{V}}(\tau) \to_P \boldsymbol{V}(\tau)$, *where* $\widehat{\boldsymbol{V}}(\tau)$ *is defined in (A.5) for the sake of presentation.*

The first result of Theorem 3.1 provides the uniform convergence rate for $\widehat{\boldsymbol{A}}(\tau)$. As a consequence, it allows us to establish a joint asymptotic distribution for the estimates of the coefficients and innovation covariance in the second result. For $\delta > 5$, the usual optimal bandwidth $h_{opt} = O\left(T^{-1/5}\right)$ satisfies the condition $\frac{T^{1-\frac{4}{\delta}} h}{\log T} \to \infty$. The third result ensures the confidence interval can be constructed practically.

Before moving on to impulse responses, we consider a practical issue — the choice of the lag p, which is usually unknown in practice and needs to be decided by the data. We select the number of lags by minimizing the next information criterion:

$$\widehat{\mathsf{p}} = \underset{1 \leq \mathsf{p} \leq \mathsf{P}}{\arg\min} \, \text{IC}(\mathsf{p}), \tag{3.7}$$

9

where $\mathrm{IC}(\mathsf{p}) = \log\{\mathrm{RSS}(\mathsf{p})\} + \mathsf{p}\cdot\chi_T$, $\mathrm{RSS}(\mathsf{p}) = \frac{1}{T}\sum_{t=1}^{T}\widehat{\eta}_{\mathsf{p},t}^{\top}\widehat{\eta}_{\mathsf{p},t}$, χ_T is the penalty term, and P is a sufficiently large fixed positive integer. The next theorem shows the validity of (3.7).

Theorem 3.2. *Let Assumptions 2, 4 and 5 hold. Suppose $\frac{T^{1-\frac{4}{\delta}}h}{\log T}\to\infty$, $\max_{t\geq 1}E(\|\epsilon_t\|^4|\mathcal{F}_{t-1}) < \infty$ a.s., $\chi_T \to 0$, and $(c_T\phi_T)^{-1}\chi_T \to \infty$, where $c_T = h^2 + \left(\frac{\log T}{Th}\right)^{1/2}$ and $\phi_T = h + \left(\frac{\log T}{Th}\right)^{1/2}$. Then $\Pr(\widehat{\mathsf{p}} = p) \to 1$ as $T \to \infty$.*

In view of the conditions on χ_T, a natural choice is

$$\chi_T = \max\left\{h^3, h\left(\frac{\log T}{Th}\right)^{1/2}, \frac{\log T}{Th}\right\}\cdot\log(1/h).$$

In the supplementary Appendix B.1, we conduct intensive simulations to examine the finite sample performance of the above information criterion.

3.2 Impulse Response Functions

We now focus on the impulse responses below, which capture the dynamic interactions among the variables of interest in a wide range of practical cases.

By Proposition 3.1, the impulse response functions of $\boldsymbol{x}_t$ is asymptotically equivalent to those of $\widetilde{x}_t$. Hence, recovering the impulse response functions requires estimating $\boldsymbol{\Psi}_j(\cdot)$'s and $\boldsymbol{\omega}(\cdot)$, which is then down to the estimation of $\boldsymbol{\Phi}(\cdot)$ and $\boldsymbol{\omega}(\cdot)$ by construction. Note further that $\boldsymbol{\Phi}(\cdot)$ is a matrix consisting of the coefficients of (3.1). The estimator of $\boldsymbol{\Phi}(\cdot)$ is intuitively defined as $\widehat{\boldsymbol{\Phi}}(\cdot)$, in which we replace $\boldsymbol{A}_j(\cdot)$ of (A.6) with the corresponding estimator obtained from (3.6). Furthermore, we require $\boldsymbol{\omega}(\tau)$ to be a lower–triangular matrix in order to fulfil the identification restriction. Thus, $\widehat{\boldsymbol{\omega}}(\tau)$ is chosen as the lower triangular matrix from the Cholesky decomposition of $\widehat{\boldsymbol{\Omega}}(\tau)$ such that $\widehat{\boldsymbol{\Omega}}(\tau) = \widehat{\boldsymbol{\omega}}(\tau)\widehat{\boldsymbol{\omega}}^{\top}(\tau)$.

With the above notation in hand, we are now ready to present the estimator of $\boldsymbol{B}_j(\tau)$ by

$$\widehat{\boldsymbol{B}}_j(\tau) = \widehat{\boldsymbol{\Psi}}_j(\tau)\widehat{\boldsymbol{\omega}}(\tau),$$

where $\widehat{\boldsymbol{\Psi}}_j(\tau) = \boldsymbol{J}\widehat{\boldsymbol{\Phi}}^j(\tau)\boldsymbol{J}^{\top}$. The corresponding asymptotic results are summarized in the following theorem.

Theorem 3.3. *Let Assumptions 2, 4 and 5 hold, and let $T \to \infty$. Suppose further that $\frac{T^{1-\frac{4}{\delta}}h}{\log T}\to\infty$, $\max_{t\geq 1}E\left(\|\epsilon_t\|^4|\mathcal{F}_{t-1}\right) < \infty$ a.s. and conditional on $\mathcal{F}_{t-1}$, the third and fourth moments of ϵ_t are identical to the corresponding unconditional moments a.s.. Then for any fixed integer $j \geq 0$*

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

where

$$\boldsymbol{\Sigma}_{\boldsymbol{B}_j}(\tau) = [\boldsymbol{C}_{j,1}(\tau), \boldsymbol{C}_{j,2}(\tau)]\boldsymbol{V}(\tau)[\boldsymbol{C}_{j,1}(\tau), \boldsymbol{C}_{j,2}(\tau)]^{\top},$$

$$\boldsymbol{B}_j^{(2)}(\tau) = \boldsymbol{C}_{j,1}(\tau)\mathrm{vec}\left(\boldsymbol{A}^{(2)}(\tau)\right) + \boldsymbol{C}_{j,2}(\tau)\mathrm{vech}\left(\boldsymbol{\Omega}^{(2)}(\tau)\right),$$

$$\boldsymbol{C}_{0,1}(\tau) = 0, \quad \boldsymbol{C}_{0,2}(\tau) = \boldsymbol{L}_d^\top \left(\boldsymbol{L}_d(\boldsymbol{I}_{d^2} + \boldsymbol{K}_{dd})(\boldsymbol{\omega}(\tau) \otimes \boldsymbol{I}_d)\boldsymbol{L}_d^\top\right)^{-1},$$

$$\boldsymbol{C}_{j,1}(\tau) = (\boldsymbol{\omega}^\top(\tau) \otimes \boldsymbol{I}_d)\left(\sum_{m=0}^{j-1} \boldsymbol{J}(\boldsymbol{\Phi}^\top(\tau))^{j-1-m} \otimes (\boldsymbol{J}\boldsymbol{\Phi}^m(\tau)\boldsymbol{J}^\top)\right) \cdot [\boldsymbol{0}_{d^2p \times d}, \boldsymbol{I}_{d^2p}], \ j \geq 1,$$

$$\boldsymbol{C}_{j,2}(\tau) = (\boldsymbol{I}_d \otimes (\boldsymbol{J}\boldsymbol{\Phi}^j(\tau)\boldsymbol{J}^\top))\boldsymbol{L}_d^\top \left(\boldsymbol{L}_d(\boldsymbol{I}_{d^2} + \boldsymbol{K}_{dd})(\boldsymbol{\omega}(\tau) \otimes \boldsymbol{I}_d)\boldsymbol{L}_d^\top\right)^{-1}, \ j \geq 1,$$

in which the elimination matrix $\boldsymbol{L}_d$ satisfies that $\mathrm{vech}(\boldsymbol{F}) = \boldsymbol{L}_d\mathrm{vec}(\boldsymbol{F})$ for any $d \times d$ matrix $\boldsymbol{F}$, and the commutation matrix $\boldsymbol{K}_{mn}$ satisfies that $\boldsymbol{K}_{mn}\mathrm{vec}(\boldsymbol{G}) = \mathrm{vec}(\boldsymbol{G}^\top)$ for any $m \times n$ matrix $\boldsymbol{G}$.

To close this section, we comment on how to construct the confidence interval. Since $\widehat{\boldsymbol{\Phi}}(\tau) \to_P \boldsymbol{\Phi}(\tau), \widehat{\boldsymbol{\omega}}(\tau) \to_P \boldsymbol{\omega}(\tau)$ and $\widehat{\boldsymbol{V}}(\tau) \to_P \boldsymbol{V}(\tau)$ by Theorem 3.1, it is straightforward to have $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{B}_j}(\tau) \to_P \boldsymbol{\Sigma}_{\boldsymbol{B}_j}(\tau)$, where $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{B}_j}(\tau)$ has a form identical to $\boldsymbol{\Sigma}_{\boldsymbol{B}_j}(\tau)$ but replacing $\boldsymbol{\Phi}(\tau), \boldsymbol{\omega}(\tau)$ and $\boldsymbol{V}(\tau)$ with their estimators, respectively.

We next show in Section 4 about how to apply the proposed model and estimation method to an empirical data. Our findings show that the estimated coefficient matrices and impulse response functions capture various time–varying features.

4 Empirical Study

In this section, we study the transmission mechanism of the monetary policy, and infer the long–run level of inflation (i.e., trend inflation) and the natural rate of unemployment (NAIRU). 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. See Stock and Watson (2016a) for more relevant discussions.

As well documented, inflation is higher and more volatile during 1970–1980, but substantially decreases in the subsequent period, which is often referred to as the Great Moderation (Primiceri, 2005). The literature has considered two main classes of explanations: bad policy or bad luck. The first type of explanations focuses on the changes in the transmission mechanism (e.g., Cogley and Sargent, 2005), while the second regards it as a consequence of changes in the size of exogenous shocks (e.g., Sims and Zha, 2006). In what follows, we revisit the arguments associated with the Great Moderation using our approach. Also, we use the VMA(∞) representation of the VAR(p) model to infer the path of trend inflation and NAIRU over time.

4.1 In–Sample Study

First, we estimate the time–varying VAR(p) model 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, representing the non–policy block, and the interest rate (measured by the average value for the Federal funds rates over the quarter), representing the monetary policy block. To isolate the monetary policy shocks, the interest rate is ordered last in the VAR model, and is treated as the monetary policy instrument. The identification requirement is that the monetary policy actions affect the inflation and the unemployment with at least one period of lag (Primiceri, 2005). The data are quarterly observations measured at an annual rate from 1954:Q3 to 2020:Q1, which are taken from the Federal Reserve Bank of St. Louis economic database. Figure 1 plots the three macro variables.

The order of the VAR(p) model and the optimal bandwidth are determined by the information criterion (3.7) and the cross validation criterion (B.1), respectively. We obtain $\widehat{\mathsf{p}} = 3$ and $\widehat{h}_{cv} = 0.435$. In the literature, the lag length is often assumed to be known with the values varying from 2 to 4, while our data–driven method indicates that 3 is the optimal value.

We first consider measuring the changes in the size of exogenous shocks. Figure 2 plots the estimated volatilities of the innovations as well as the associated 95% confidence intervals. The figure exhibits evidence for a general decline in unconditional volatilities. Our results thus support the "bad luck" explanations (e.g., Primiceri, 2005; Sims and Zha, 2006).

We then consider the responses of the inflation to the monetary policy shocks. Figure 3 plots the time–varying impulse responses of the inflation to a structural monetary shock as well as the associated 95% confidence intervals. It is clear that the confidence intervals are much wider at the beginning of the sample period implying higher uncertainty before 1970. On the other hand, the structural responses of the inflation seem to be statistically insignificant from 1970 to 2010. Thus, we conclude that the monetary shocks have less influence on the inflation before and during the period of the Great Moderation.

Finally, 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 $\boldsymbol{\mu}_t = \lim_{p \to \infty} E_t(\boldsymbol{x}_{t+p}) = (\boldsymbol{I} - \boldsymbol{A}_t)^{-1}\boldsymbol{a}_t$, where $\boldsymbol{a}_t$ is the intercept term and $\boldsymbol{A}_t$ is the autoregressive coefficients. The main difference between our method and Petrova's method is that we invert time–varying VAR(p) model to the time–varying MA(∞) model, and then explicitly estimate the underlying trends of the inflation and the NAIRU using model (2.7).

Figure 4 plots the estimates of the trend inflation and the NAIRU. It is obvious that the underlying trend of the inflation is high in the 1970s, but decreases in the subsequent period. After the Great Moderation, the long–run level of the 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.

4.2 Out–of–Sample Forecasting

In this subsection, we focus on the out–of–sample forecasting, and compare the forecasts of a Bayesian time–varying parameter VAR with stochastic volatility (TVP–SV) (cf., Primiceri, 2005), as well as a VAR model with constant parameters (CVAR).

Specifically, we consider the 1–8 quarters ahead forecasts. That is to forecast

$$\overline{\boldsymbol{x}}_{t+1|t+h} = h^{-1} \sum_{i=1}^{h} \boldsymbol{x}_{t+i}$$

for $h = 1$, 2, 4 and 8, where $\boldsymbol{x}_t$ includes the values of the three aforementioned macro variables at the date t. The forecasts of $\overline{\boldsymbol{x}}_{t+1|t+h}$ are constructed by $\widehat{\overline{\boldsymbol{x}}}_{t+1|t+h} = \widehat{\boldsymbol{A}}_t \boldsymbol{z}_t$, in which $\widehat{\boldsymbol{A}}_t$ is estimated from CVAR, TVP–SV and the time–varying VAR model using the available data at time t and $\boldsymbol{z}_t = [1, \boldsymbol{x}_t^\top, \boldsymbol{x}_{t-1}^\top, \boldsymbol{x}_{t-2}^\top]^\top$. The expanding window scheme is adopted. For comparison, we compute the root of mean square errors (RMSE) for CVAR as a benchmark, and the RMSE ratios for others. The out–of–sample forecast period covers 1985:Q2–end, about 35 years.

The forecasting results are presented in Table 1, in which the values represent the ratios of the RMSEs of the corresponding method over the RMSEs of the benchmark method (i.e., CVAR). The result shows that the time–varying VAR model and TVP–SV perform much better than CVAR, which implies the desirability of introducing variations in forecasting models. In addition, the time–varying VAR model has a better forecasting performance than TVP–SV with the increase of the forecast horizon.

Table 1: The 1–8 quarters ahead forecast. The values represent the ratios of the RMSEs of the corresponding method over the RMSEs of the benchmark method (i.e., CVAR). In each panel, the numbers in bold font of each column indicate the method that provides the best out–of–sample forecast for given h.

	$h = 1Q$	$h = 2Q$	$h = 4Q$	$h = 8Q$
	Inflation rate, 1985:Q1–end			
TVP–SV	**0.9771**	0.9841	0.9850	0.9880
time–varying VAR	0.9986	**0.9732**	**0.9656**	**0.9855**
	Unemployment rate, 1985:Q1–end			
TVP–SV	**0.9119**	0.9683	1.0228	1.0519
time–varying VAR	0.9344	**0.9671**	**0.9928**	1.0076
	Interest rate, 1985:Q1–end			
TVP–SV	**0.8388**	**0.8773**	**0.9435**	1.0066
time–varying VAR	0.9720	0.9743	0.9738	**0.9892**
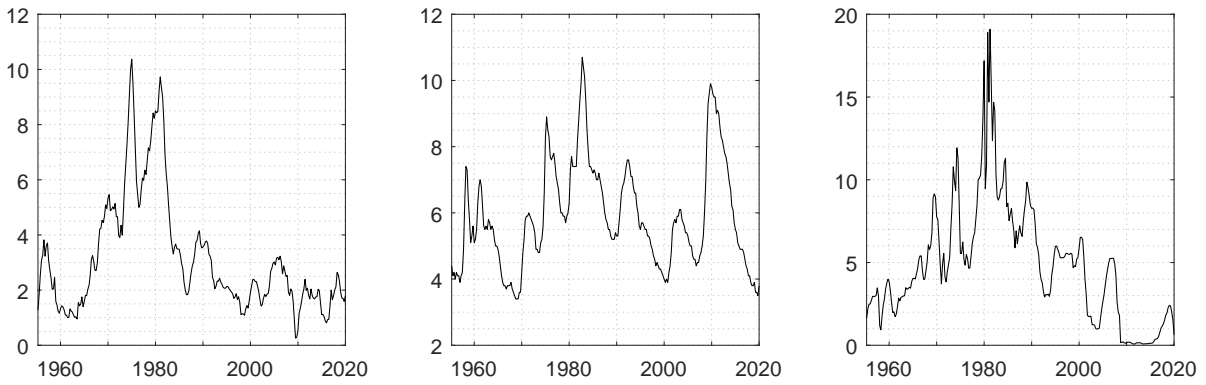
Figure 1: Plots of the inflation (left), the unemployment rate (middle) and the interest rate (right)
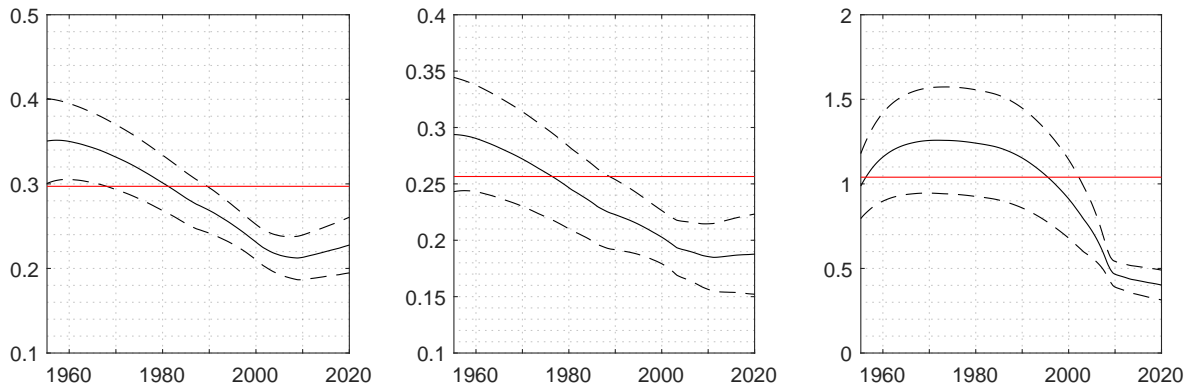
Figure 2: The estimated volatilities of the innovations in the inflation equation (left), the unemployment equation (middle) and the interest rate equation (right) as well as the associated 95% confidence intervals. The red line denotes the estimated volatilities using the constant VAR model.

Figure 3: The time–varying structural impulse response of the inflation to monetary policy shocks as well as the associated 95% confidence interval.

Figure 4: The estimated trends of the inflation and the NAIRU as well as the associated 95% confidence intervals
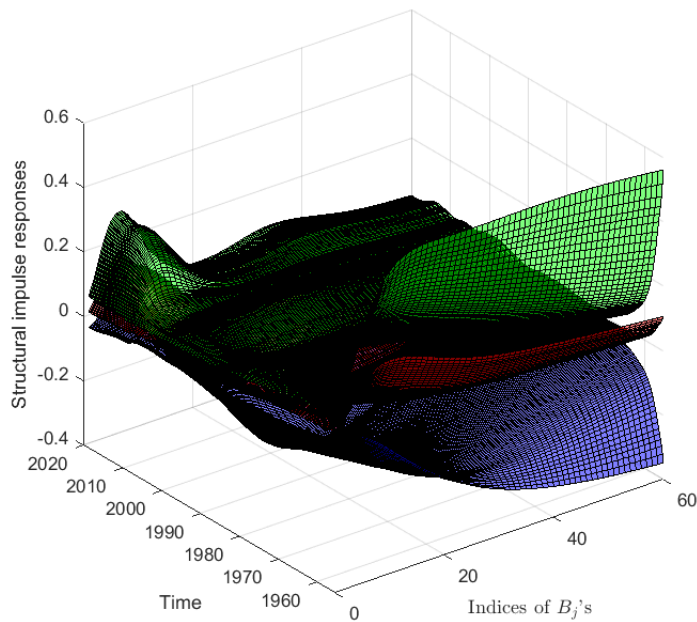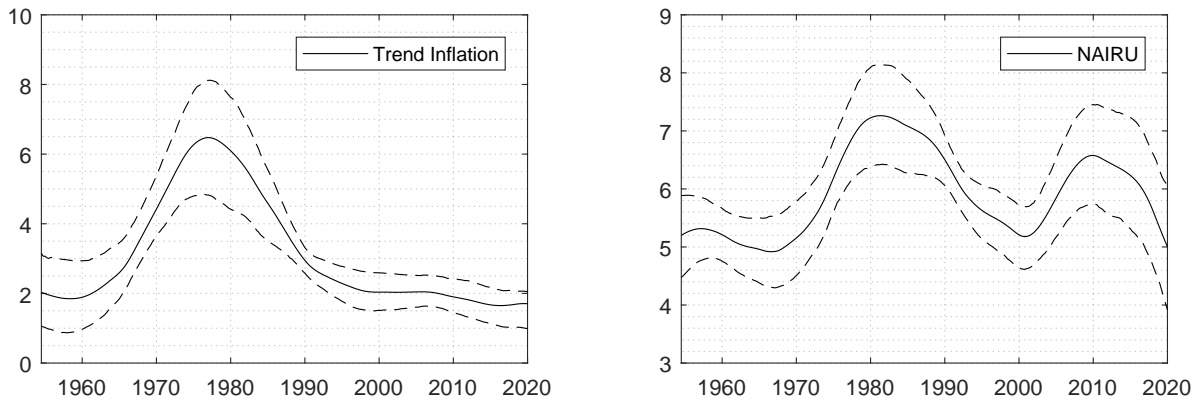
5 Conclusion

This paper has proposed a class of time–varying VMA(∞) models, which nest, for instance, time–varying VAR, time–varying VARMA, and so forth as special cases. Both the estimation methodology and asymptotic theory have been established accordingly. In the empirical study, we have investigated the transmission mechanism of monetary policy using U.S. data, and uncover a fall in the volatilities of exogenous shocks. Our findings include (i) monetary policy shocks have less influence on inflation before and during the so–called Great Moderation, (ii) inflation is more anchored recently, and (iii) the long–run level of inflation is below, but quite close to the Federal Reserve's target of two percent after the beginning of the Great Moderation period. In addition, in Appendix B of the online supplementary material, we have evaluated the finite–sample performance of the proposed model and estimation theory.

There are several directions for possible extensions. The first one is to test about whether the d–dimensional components of the VAR(p) process is cross–sectionally independent for the case where the dimensionality, d, and the number of lags, p, may diverge along with the sample size, T. The second one is about model specification testing to check whether some of the time–varying coefficient matrices, $\boldsymbol{A}_j(\tau)$, may just be constant matrices, $\boldsymbol{A}_{j0}$. Existing studies by Gao and Gijbels (2008), Pan et al. (2014), and Chen and Wu (2019) may be useful for both issues. The third one is to allow for some cointegrated structure in our settings. The recent work by Zhang et al. (2019) provides us with a good reference. We wish to leave such issues for future study.

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

Cai, Z. (2007), 'Trending time–varying coefficient time series models with serially correlated errors', *Journal of Econometrics* **136**(2), 163–188.

Chen, L. and Wu, W. B. (2019), 'Testing for trends in high-dimensional time series', *Journal of the American Statistical Association* **114**(526), 869–881.

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.

Cogley, T. and Sargent, T. J. (2005), 'Drifts and volatilities: Monetary policies and outcomes in the post World War II U.S.', *Review of Economic Dynamics* **8**(2), 262–302.

Coibion, O. and Gorodnichenko, Y. (2011), 'Monetary policy, trend inflation, and the great moderation: An alternative interpretation', *American Economic Review* **101**(1), 341–70.

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., Neumann, M. H. and Von Sachs, R. (1999), 'Nonlinear wavelet estimation of time-varying autoregressive processes', *Bernoulli* **5**(5), 873–906.

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.

Dahlhaus, R., Richter, S. and Wu, W. B. (2019), 'Towards a general theory for nonlinear locally stationary processes', *Bernoulli* **25**(2), 1013–1044.

Fan, J. and Yao, Q. (2003), *Nonlinear Time Series: Parametric and Nonparametric Methods*, Springer–Verlag, New York.

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. (2007), *Nonlinear Time Series: Semi– and Non–Parametric Methods*, Chapman & Hall/CRC, London.

Gao, J. and Gijbels, I. (2008), 'Bandwidth selection in nonparametric kernel testing', *Journal of the American Statistical Association* **103**(484), 1584–1594.

Giannone, D., Lenza, M. and Primiceri, G. E. (2019), 'Priors for the long run', *Journal of the American Statistical Association* **114**(526), 565–580.

Giraitis, L., Kapetanios, G. and Yates, T. (2014), 'Inference on stochastic time-varying coefficient models', *Journal of Econometrics* **179**(1), 46–65.

Giraitis, L., Kapetanios, G. and Yates, T. (2018), 'Inference on multivariate heteroscedastic time varying random coefficient models', *Journal of Time Series Analysis* **39**(2), 129–149.

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. (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.

Härdle, W., Liang, H. and Gao, J. (2000), *Partially Linear Models*, Springer-Verlag, New York.

Justiniano, A. and Primiceri, G. E. (2008), 'The time-varying volatility of macroeconomic fluctuations', *American Economic Review* **98**(3), 604–41.

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

Li, D., Phillips, P. C. and Gao, J. (2019), 'Kernel-based inference in time-varying coefficient cointegrating regression', *Journal of Econometrics* **215**(2), 607–632.

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.

Pan, G., Gao, J. and Yang, Y. (2014), 'Testing independence for a large number of high dimensional random vectors', *Journal of the American Statistical Association* **109**(506), 600–612.

Paul, P. (2019), 'The time-varying effect of monetary policy on asset prices', *Review of Economics and Statistics* p. forthcoming.

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

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

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

Robinson, P. M. (1989), 'Chapter 15: Nonparametric estimation of time-varying parameters', *Statistical Analysis and Forecasting of Economic Structural Change* pp. 253–264.

Rohan, N. and Ramanathan, T. (2013), 'Nonparametric estimation of a time-varying GARCH model', *Journal of Nonparametric Statistics* **25**(1), 33–52.

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

Sims, C. A. (1980), 'Macroeconomics and reality', *Econometrica* **48**(1), 1–48.

Sims, C. and Zha, T. (2006), 'Were there regime switches in U.S. monetary policy?', *American Economic Review* **96**(1), 54–81.

Stock, J. H. and Watson, M. W. (2001), 'Vector autoregressions', *Journal of Economic Perspectives* **15**(4), 101–115.

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

Stock, J. H. and Watson, M. W. (2016*b*), Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics, *in* 'Handbook of macroeconomics', Vol. 2, Elsevier, pp. 415–525.

Truquet, L. (2017), 'Parameter stability and semiparametric inference in time varying auto-regressive conditional heteroscedasticity models', *Journal of the Royal Statistical Society: Series B* **79**(5), 1391–1414.

Tsay, R. S. (1998), 'Testing and modeling multivariate threshold models', *Journal of the American Statistical Association* **93**(443), 1188–1202.

Vogt, M. (2012), 'Nonparametric regression for locally stationary time series', *Annals of Statistics* **40**(5), 2601–2633.

Wang, Q. and Chan, N. (2014), 'Uniform convergence rates for a class of martingales with application in non-linear cointegrating regression', *Bernoulli* **20**(1), 207–230.

Wu, W. B. and Zhao, Z. (2007), 'Inference of trends in time series', *Journal of the Royal Statistical Society: Series B* **69**(3), 391–410.

Zhang, R., Robinson, P. and Yao, Q. (2019), 'Identifying cointegration by eigenanalysis', *Journal of the American Statistical Association* **114**(526), 916–927.

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

Appendix A

For the sake of presentation, we first provide some notation and mathematical sybmols in Appendix A.1, and then present the dependent wild bootstrap (DWB) procedure with the associated asymptotic properties in Appendix A.2. Some proofs of the main results are provided in Appendix A.3. Simulations, some secondary results and omitted proofs are given in the online supplementary Appendix B. In what follows, M and $O(1)$ always stand for constants, and may be different at each appearance.

A.1 Notation and Mathematical Symbols

For ease of notation, we define three matrices $\boldsymbol{\Sigma}(\tau)$, $\boldsymbol{V}(\tau)$ and $\boldsymbol{\Phi}(\tau)$ with their estimators respectively. First, for $\forall \tau \in (0,1)$, let

$$\boldsymbol{\Sigma}(\tau) = \begin{pmatrix} 1 & \boldsymbol{\mu}^\top(\tau) & \cdots & \boldsymbol{\mu}^\top(\tau) \\ \boldsymbol{\mu}(\tau) & \boldsymbol{\Sigma}_0(\tau) & \cdots & \boldsymbol{\Sigma}_{p-1}^\top(\tau) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mu}(\tau) & \boldsymbol{\Sigma}_{p-1}(\tau) & \cdots & \boldsymbol{\Sigma}_0(\tau) \end{pmatrix}, \tag{A.1}$$

in which $\boldsymbol{\mu}(\tau)$ and $\boldsymbol{B}_j(\tau)$ are defined in Proposition 3.1 and $\boldsymbol{\Sigma}_m(\tau) = \boldsymbol{\mu}(\tau)\boldsymbol{\mu}(\tau)^\top + \sum_{j=0}^\infty \boldsymbol{B}_j(\tau)\boldsymbol{B}_{j+m}^\top(\tau)$ for $m = 0, \ldots, p-1$. We define the estimator of $\boldsymbol{\Sigma}(\tau)$ as

$$\widehat{\boldsymbol{\Sigma}}(\tau) = \frac{1}{T} \sum_{t=1}^T \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-1}^\top K_h(\tau_t - \tau), \tag{A.2}$$

where $\boldsymbol{z}_t$ is defined in (3.5).

Next, we let

$$\boldsymbol{V}(\tau) = \begin{pmatrix} \boldsymbol{V}_{1,1}(\tau) & \boldsymbol{V}_{2,1}^\top(\tau) \\ \boldsymbol{V}_{2,1}(\tau) & \boldsymbol{V}_{2,2}(\tau) \end{pmatrix}, \tag{A.3}$$

where

$$\boldsymbol{V}_{1,1}(\tau) = \tilde{v}_0 \boldsymbol{\Sigma}^{-1}(\tau) \otimes \boldsymbol{\Omega}(\tau),$$

$$\boldsymbol{V}_{2,1}(\tau) = \lim_{T\to\infty} \frac{h}{T} \sum_{t=1}^T E\left(\text{vech}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top) \boldsymbol{\eta}_t^\top \boldsymbol{Z}_{t-1}^\top \right) K_h(\tau_t - \tau)^2 \cdot (\boldsymbol{\Sigma}^{-1}(\tau) \otimes \boldsymbol{I}_d),$$

$$\boldsymbol{V}_{2,2}(\tau) = \lim_{T\to\infty} \frac{h}{T} \sum_{t=1}^T E\left(\text{vech}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top) \text{vech}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top)^\top \right) K_h(\tau_t - \tau)^2$$

$$- \tilde{v}_0 \text{vech}\left(\boldsymbol{\Omega}(\tau)\right) \text{vech}\left(\boldsymbol{\Omega}(\tau)\right)^\top. \tag{A.4}$$

The estimator of $\boldsymbol{V}(\tau)$ is then defined as follows.

$$\widehat{\boldsymbol{V}}(\tau) = \begin{pmatrix} \widehat{\boldsymbol{V}}_{1,1}(\tau) & \widehat{\boldsymbol{V}}_{2,1}^\top(\tau) \\ \widehat{\boldsymbol{V}}_{2,1}(\tau) & \widehat{\boldsymbol{V}}_{2,2}(\tau) \end{pmatrix}, \tag{A.5}$$

where $\widehat{\boldsymbol{V}}_{1,1}(\tau)$, $\widehat{\boldsymbol{V}}_{2,1}(\tau)$ and $\widehat{\boldsymbol{V}}_{2,2}(\tau)$ have the forms identical to their counterparts of (A.4), but we replace $\boldsymbol{\Sigma}(\tau)$, $\boldsymbol{\eta}_t$ and $\boldsymbol{\Omega}(\tau)$ with their estimators presented in (A.2) and (3.6).

Finally, recall the following definition:

$$
\boldsymbol{\Phi}(\tau) = \begin{pmatrix} \boldsymbol{A}_1(\tau) & \cdots & \boldsymbol{A}_{p-1}(\tau) & \boldsymbol{A}_p(\tau) \\ \boldsymbol{I}_d & \cdots & \boldsymbol{0}_d & \boldsymbol{0}_d \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0}_d & \cdots & \boldsymbol{I}_d & \boldsymbol{0}_d \end{pmatrix}. \tag{A.6}
$$

Replacing $\boldsymbol{A}_j(\tau)$'s of (A.6) with their estimators obtained from (3.6) yields the estimator $\widehat{\boldsymbol{\Phi}}(\tau)$ straight away.

A.2 Dependent Wild Bootstrap

We now present the detailed dependent wild bootstrap (DWB) procedure which is used to establish the confidence interval associated with Theorem 2.2.

1. For $\forall \tau \in (0,1)$, let $\widetilde{\boldsymbol{\mu}}(\tau)$ be the same as defined in (2.8) using an over–smoothing bandwidth $\widetilde{h}$. Obtain residuals $\widehat{\boldsymbol{e}}_t = \boldsymbol{x}_t - \widetilde{\boldsymbol{\mu}}(\tau_t)$ for $t \geq 1$.

2. Generate the bootstrap observations $\boldsymbol{x}_t^* = \widetilde{\boldsymbol{\mu}}(\tau_t) + \boldsymbol{e}_t^*$ for $t \geq 1$, where $\boldsymbol{e}_t^* = \xi_t^* \widehat{\boldsymbol{e}}_t$, $\{\xi_t^*\}$ is an l-dependent process satisfying $E(\xi_t^*) = 0$, $E|\xi_t^*|^2 = 1$, $E|\xi_t^*|^\delta < \infty$ with $\delta > 2$, and $E(\xi_t^* \xi_s^*) = a\left((t-s)/l\right)$ with some kernel function $a\left(\cdot\right)$ and tuning parameter l.

3. For $\forall \tau \in (0,1)$, let $\widehat{\boldsymbol{\mu}}^*(\tau)$ be the same as (2.8) but using $\{\boldsymbol{x}_t^*\}$.

4. Repeat Steps 2–3 J times. Let $\boldsymbol{q}_\alpha(\tau)$ be the α–quantile of 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 $\left[\widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{q}_{1-\alpha/2}(\tau), \widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{q}_{\alpha/2}(\tau)\right]$.

The above DWB procedure requires a tuning parameter l, which is the so–called "block length" (Shao, 2010). The following conditions are required to ensure the validity of the DWB procedure.

Assumption A.1. *Suppose that $l \to \infty$, $\max\{\widetilde{h}, h/\widetilde{h}\} \to 0$ and $l \cdot \max\{1/\sqrt{Th}, \widetilde{h}^4, 1/T\widetilde{h}\} \to 0$. Additionally, let $a(\cdot)$ be a symmetric kernel defined on $[-1,1]$ satisfying that $a(0) = 1$ and $a(\cdot)$ is continuous at 0 with $a^{(1)}(0) < \infty$.*

We summarize the asymptotic properties of the DWB method by the next theorem.

Theorem A.1. *Let Assumptions 2–4 and A.1 hold. For $\forall \tau \in (0,1)$, as $T \to \infty$,*

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.

We comment on some practical issues.

Bandwidth Selection: Since the observations are dependent, 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 h_{mcv} is selected to minimize the following objective function:

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

where $\widehat{\boldsymbol{\mu}}_{k,h}(\tau) = \left(\sum_{t:|t-\tau T|>k} K\left(\frac{\tau_t-\tau}{h}\right)\right)^{-1} \sum_{t:|t-\tau T|>k} \boldsymbol{x}_t K\left(\frac{\tau_t-\tau}{h}\right)$. Moreover, in the first step of the bootstrap procedure, we follow the suggestion of Bühlmann et al. (1998) to use $\widetilde{h} = c_0 \cdot h_{mcv}^{5/9}$ with $c_0 = 2$.

Tuning parameter: In the second step of the bootstrap procedure, we choose the kernel function $a(\cdot)$ and the bandwidth l as in Shao (2010).

A.3   Proofs of the Main Results

Proof of Theorem 2.1.

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

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

where $\mathbb{B}_t(1)$ and $\widetilde{\mathbb{B}}_t(L)$ have been defined in the first result of Lemma B.3. We are then able to write

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

$$
\leq \sup_{\tau \in [a,b]} \left\| \sum_{t=1}^{T} \boldsymbol{W}_{T,t}(\tau)\mathbb{B}_t(1)\boldsymbol{\epsilon}_t \right\| + \sup_{\tau \in [a,b]} \left\| \boldsymbol{W}_{T,1}(\tau)\widetilde{\mathbb{B}}_1(L)\boldsymbol{\epsilon}_0 \right\| + \sup_{\tau \in [a,b]} \left\| \boldsymbol{W}_{T,T}(\tau)\widetilde{\mathbb{B}}_T(L)\boldsymbol{\epsilon}_T \right\|
$$

$$
+ \sup_{\tau \in [a,b]} \left\| \sum_{t=1}^{T-1} \left(\boldsymbol{W}_{T,t+1}(\tau)\widetilde{\mathbb{B}}_{t+1}(L) - \boldsymbol{W}_{T,t}(\tau)\widetilde{\mathbb{B}}_t(L) \right) \boldsymbol{\epsilon}_t \right\|
$$

$$
:= I_{T,1} + I_{T,2} + I_{T,3} + I_{T,4},
$$

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

By Lemma B.5, 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 B.3. Thus, we just need 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 B.3;

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

(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

$$\sup_{\tau \in [a,b]} \left\| \sum_{t=1}^{T-1} \left(\boldsymbol{W}_{T,t+1}(\tau)\widetilde{\mathbb{B}}_{t+1}(L) - \boldsymbol{W}_{T,t}(\tau)\widetilde{\mathbb{B}}_t(L) \right) \boldsymbol{\epsilon}_t \right\|$$

$$= \sup_{\tau \in [a,b]} \left\| \sum_{t=1}^{T-1} (\boldsymbol{W}_{T,t+1}(\tau) - \boldsymbol{W}_{T,t}(\tau))\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\epsilon}_t + \boldsymbol{W}_{T,t}(\tau)(\widetilde{\mathbb{B}}_{t+1}(L) - \widetilde{\mathbb{B}}_t(L))\boldsymbol{\epsilon}_t \right\|$$

$$\leq \max_{1 \leq t \leq T-1} \|\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\epsilon}_t\| \cdot \sup_{\tau \in [a,b]} \sum_{t=1}^{T-1} \|\boldsymbol{W}_{T,t+1}(\tau) - \boldsymbol{W}_{T,t}(\tau)\|$$

$$+ \sup_{\tau \in [a,b], 1 \leq t \leq T} \|\boldsymbol{W}_{T,t}(\tau)\| \cdot \sum_{t=1}^{T-1} \|(\widetilde{\mathbb{B}}_{t+1}(L) - \widetilde{\mathbb{B}}_t(L))\boldsymbol{\epsilon}_t\|$$

$$= O_P(T^{1/\delta} \cdot d_T) + O_P(d_T) = o_P(\sqrt{d_T \log T}).$$

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.

$$\sup_{\tau \in [a,b]} \left\| \sum_{t=1}^{T} \text{vec} \left(\boldsymbol{W}_{T,t}(\tau)(\boldsymbol{x}_t \boldsymbol{x}_t^\top - E(\boldsymbol{x}_t \boldsymbol{x}_t^\top)) \right) \right\|$$

$$\leq 2 \sup_{\tau \in [a,b]} \left\| \sum_{t=1}^{T} (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(\tau)) \sum_{j=0}^{\infty} (\boldsymbol{B}_{j,t} \otimes \boldsymbol{\mu}_t)\boldsymbol{\epsilon}_{t-j} \right\|$$

$$+ \sup_{\tau \in [a,b]} \left\| \sum_{t=1}^{T} (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(\tau)) \sum_{j=0}^{\infty} (\boldsymbol{B}_{j,t} \otimes \boldsymbol{B}_{j,t}) \left(\text{vec} \left(\boldsymbol{\epsilon}_{t-j} \boldsymbol{\epsilon}_{t-j}^\top \right) - \text{vec} (\boldsymbol{I}_d) \right) \right\|$$

$$+ 2 \sup_{\tau \in [a,b]} \left\| \sum_{t=1}^{T} (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(\tau)) \sum_{r=1}^{\infty} \sum_{j=0}^{\infty} (\boldsymbol{B}_{j+r,t} \otimes \boldsymbol{B}_{j,t})\text{vec} (\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-j-r}) \right\|$$

$$:= 2I_{T,1} + I_{T,2} + 2I_{T,3},$$

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

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

$$I_{T,2} \leq \sup_{\tau \in [a,b]} \left\| \sum_{t=1}^{T} (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(\tau))\mathbb{B}_t^0(1) \left(\text{vec} \left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top \right) - \text{vec} (\boldsymbol{I}_d) \right) \right\|$$

$$+ \sup_{\tau \in [a,b]} \left\| (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,1}(\tau))\widetilde{\mathbb{B}}_1^0(L)\text{vec} \left(\boldsymbol{\epsilon}_0 \boldsymbol{\epsilon}_0^\top \right) \right\| + \sup_{\tau \in [a,b]} \left\| (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,T}(\tau))\widetilde{\mathbb{B}}_T^0(L)\text{vec} \left(\boldsymbol{\epsilon}_T \boldsymbol{\epsilon}_T^\top \right) \right\|$$

$$+ \sup_{\tau \in [a,b]} \left\| \sum_{t=1}^{T-1} \left((\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t+1}(\tau))\widetilde{\mathbb{B}}_{t+1}^0(\widetilde{L}) - (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(\tau))\widetilde{\mathbb{B}}_t^0(L) \right) \cdot \text{vec} \left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top \right) \right\|$$

$$:= I_{T,21} + I_{T,22} + I_{T,23} + I_{T,24}.$$

By Lemma B.6, 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 B.3. Similar to the proof of the first result, for $I_{T,24}$, we write

$$\sup_{\tau \in [a,b]} \left\| \sum_{t=1}^{T-1} \left((\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t+1}(\tau)) \widetilde{\mathbb{B}}_{t+1}^0(L) - (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(\tau)) \widetilde{\mathbb{B}}_t^0(L) \right) \operatorname{vec}\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top\right) \right\|$$

$$\leq \sqrt{d} \sup_{\tau \in [a,b], 1 \leq t \leq T} \|\boldsymbol{W}_{T,t+1}(\tau)\| \cdot \sum_{t=1}^{T-1} \left\| \left(\widetilde{\mathbb{B}}_{t+1}^0(L) - \widetilde{\mathbb{B}}_t^0(L) \right) \operatorname{vec}\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top\right) \right\|$$

$$+ \sqrt{d} \max_t \left\| \widetilde{\mathbb{B}}_t^0(L) \operatorname{vec}\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top\right) \right\| \cdot \sup_{\tau \in [a,b]} \sum_{t=1}^{T-1} \|\boldsymbol{W}_{T,t+1}(\tau) - \boldsymbol{W}_{T,t}(\tau)\|$$

$$= o_P\left(\sqrt{d_T \log T}\right),$$

where we have used the facts that

(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). $\sum_{t=1}^{T-1} \left\| \widetilde{\mathbb{B}}_{t+1}^0(1) - \widetilde{\mathbb{B}}_t^0(1) \right\| = O(1)$.

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 B.3, we have

$$\sup_{\tau \in [a,b]} \left\| \sum_{t=1}^T (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(\tau)) \sum_{r=1}^\infty \sum_{j=0}^\infty (\boldsymbol{B}_{j+r,t} \otimes \boldsymbol{B}_{j,t}) \operatorname{vec}\left(\boldsymbol{\epsilon}_{t-j} \boldsymbol{\epsilon}_{t-j-r}\right) \right\|$$

$$\leq \sup_{\tau \in [a,b]} \left\| \sum_{t=1}^T (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(\tau)) \boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t \right\| + \sup_{\tau \in [a,b]} \left\| (\boldsymbol{I}_d \otimes W_{T,1}(\tau)) \sum_{r=1}^\infty \widetilde{\mathbb{B}}_1^r(L) \operatorname{vec}\left(\boldsymbol{\epsilon}_0 \boldsymbol{\epsilon}_{-r}^\top\right) \right\|$$

$$+ \sup_{\tau \in [a,b]} \left\| (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,T}(\tau)) \sum_{r=1}^\infty \widetilde{\mathbb{B}}_T^r(L) \operatorname{vec}\left(\boldsymbol{\epsilon}_T \boldsymbol{\epsilon}_{T-r}^\top\right) \right\|$$

$$+ \sup_{\tau \in [a,b]} \left\| \sum_{t=1}^{T-1} \sum_{r=1}^\infty \left((\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t+1}(\tau)) \widetilde{\mathbb{B}}_{t+1}^r(L) - (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(\tau)) \widetilde{\mathbb{B}}_t^r(L) \right) \operatorname{vec}\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_{t-r}^\top\right) \right\|$$

$$:= I_{T,31} + I_{T,32} + I_{T,33} + I_{T,34},$$

where $\boldsymbol{\zeta}_t$ is defined in Lemma B.6.

By Lemma B.6, $I_{T,31} = O_P\left(\sqrt{d_T \log T}\right)$. 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 B.3. For $I_{T,34}$, we write

$$\sup_{\tau \in [a,b]} \left\| \sum_{t=1}^{T-1} \sum_{r=1}^\infty \left((\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t+1}(\tau)) \widetilde{\mathbb{B}}_{t+1}^r(L) - (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(\tau)) \widetilde{\mathbb{B}}_t^r(L) \right) \operatorname{vec}\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_{t-r}^\top\right) \right\|$$

$$\leq \sqrt{d} \sup_{\tau \in [a,b], 1 \leq t \leq T} \|\boldsymbol{W}_{T,t}(\tau)\| \cdot \sum_{t=1}^{T-1} \left\| \sum_{r=1}^\infty \left(\widetilde{\mathbb{B}}_{t+1}^r(L) - \widetilde{\mathbb{B}}_t^r(L) \right) \operatorname{vec}\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_{t-r}^\top\right) \right\|$$

$$+ \sqrt{d} \max_t \left\| \sum_{r=1}^\infty \widetilde{\mathbb{B}}_t^r(L) \operatorname{vec}\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_{t-r}^\top\right) \right\| \cdot \sup_{\tau \in [a,b]} \sum_{t=1}^{T-1} \|\boldsymbol{W}_{T,t+1}(\tau) - \boldsymbol{W}_{T,t}(\tau)\|$$

$$= o_P(\sqrt{d_T \log T}).$$

where we have used the fact that

(1). $T^{4/\delta} d_T \log T \to 0$;

(2). $\displaystyle\max_t \left\| \sum_{r=1}^{\infty} \widetilde{\mathbb{B}}_t^r(L) \mathrm{vec}\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_{t-r}^\top\right) \right\| = O_P(T^{2/\delta})$;

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

(4). $\displaystyle\sup_{\tau \in [a,b]} \sum_{t=1}^{T-1} \|\boldsymbol{W}_{T,t+1}(\tau) - \boldsymbol{W}_{T,t}(\tau)\| = O(d_T)$.

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

Proof of Theorem 2.2.

Write

$$\frac{1}{\sqrt{Th}} \sum_{t=1}^{T} (\boldsymbol{x}_t - \boldsymbol{\mu}(\tau_t)) K\left(\frac{\tau_t - \tau}{h}\right)$$

$$= \frac{1}{\sqrt{Th}} \sum_{t=1}^{T} \mathbb{B}_t(1) \boldsymbol{\epsilon}_t K\left(\frac{\tau_t - \tau}{h}\right) + \frac{1}{\sqrt{Th}} \widetilde{\mathbb{B}}_t(L) \boldsymbol{\epsilon}_0 K\left(\frac{\tau_1 - \tau}{h}\right) - \frac{1}{\sqrt{Th}} \widetilde{\mathbb{B}}_T(L) \boldsymbol{\epsilon}_T K\left(\frac{\tau_T - \tau}{h}\right)$$

$$+ \frac{1}{\sqrt{Th}} \sum_{t=1}^{T-1} \left(\widetilde{\mathbb{B}}_{t+1}(L) K\left(\frac{\tau_{t+1} - \tau}{h}\right) - \widetilde{\mathbb{B}}_t(L) K\left(\frac{\tau_t - \tau}{h}\right) \right) \boldsymbol{\epsilon}_t$$

$$= \frac{1}{\sqrt{Th}} \sum_{t=1}^{T} \mathbb{B}_t(1) \boldsymbol{\epsilon}_t K\left(\frac{\tau_t - \tau}{h}\right) + o_P(1),$$

where the second equality follows from the similar arguments as in the proof of Theorem 2.1.

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

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

Since

$$\mathrm{Var}\left(\frac{1}{\sqrt{Th}} \sum_{t=1}^{T} \mathbb{B}_t(1) \boldsymbol{\epsilon}_t K\left(\frac{\tau_t - \tau}{h}\right) \right) = \frac{1}{Th} \sum_{t=1}^{T} \mathbb{B}_t(1) \mathbb{B}_t^\top(1) K\left(\frac{\tau_t - \tau}{h}\right)^2$$

$$\to \widetilde{v}_0 \left\{ \sum_{j=0}^{\infty} \boldsymbol{B}_j(\tau) \right\} \left\{ \sum_{j=0}^{\infty} \boldsymbol{B}_j^\top(\tau) \right\},$$

we then use the Cramér-Wold device to prove its asymptotic normality. That is to show that for any

conformable vector $\boldsymbol{d}$

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

Let $\boldsymbol{Z}_t(\tau) = \frac{1}{\sqrt{Th}}\boldsymbol{d}^{\top}\mathbb{B}_t(1)\boldsymbol{\epsilon}_t K\left(\frac{\tau_t-\tau}{h}\right)$. By the law of large numbers of m.d.s. and the assumption $E\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^{\top}|\mathcal{F}_{t-1}\right) = \boldsymbol{I}_d$ a.s., we have for any $\tau \in (0,1)$

$$\sum_{t=1}^{T}\boldsymbol{Z}_t^2(\tau) \to_P \widetilde{v}_0\boldsymbol{d}^{\top}\left\{\sum_{j=0}^{\infty}\boldsymbol{B}_j(\tau)\right\}\left\{\sum_{j=0}^{\infty}\boldsymbol{B}_j^{\top}(\tau)\right\}\boldsymbol{d}.$$

Furthermore, for any $\nu > 0$ and $\tau \in (0,1)$, by Hölder's inequality and Markov's inequality,

$$\sum_{t=1}^{T} E\left(\boldsymbol{Z}_t^2(\tau)I\left(|\boldsymbol{Z}_t(\tau)| > \nu\right)\right)$$

$$\leq \frac{1}{Th}\sum_{t=1}^{T}K^2\left(\frac{\tau_t-\tau}{h}\right)\left(E|\boldsymbol{d}^{\top}\mathbb{B}_t(1)\boldsymbol{\epsilon}_t|^{\delta}\right)^{2/\delta}\left(\frac{E|\boldsymbol{d}^{\top}\mathbb{B}_t(1)\boldsymbol{\epsilon}_t|^{\delta}}{(Th)^{\delta/2}\nu^{\delta}}\right)^{(\delta-2)/\delta}$$

$$= O\left(\frac{1}{(Th)^{\delta-2/2}}\right) = o(1).$$

By Lemma B.1, the proof is now completed. $\qquad\square$

Proof of Theorem 3.1.

(1). By Proposition 3.1,

$$\sum_{t=1}^{T} E\left\|\boldsymbol{x}_t\boldsymbol{x}_t^{\top} - \widetilde{\boldsymbol{x}}_t\widetilde{\boldsymbol{x}}_t^{\top}\right\| \leq \sum_{t=1}^{T} E\left\|\boldsymbol{x}_t - \widetilde{\boldsymbol{x}}_t\right\| \cdot \left(E\left\|\boldsymbol{x}_t\right\| + E\left\|\widetilde{\boldsymbol{x}}_t\right\|\right) = O(1).$$

Therefore, Lemma B.7 are still valid for the time-varying VAR process (3.1). For example, consider the uniform convergence results, by Lemma B.7,

$$\sup_{\tau\in[0,1]}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{x}_t\boldsymbol{x}_t - E(\widetilde{\boldsymbol{x}}_t\widetilde{\boldsymbol{x}}_t^{\top})\right)K_h\left(\tau_t-\tau\right)\right\|$$

$$\leq \sup_{\tau\in[0,1]}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\widetilde{\boldsymbol{x}}_t\widetilde{\boldsymbol{x}}_t^{\top} - E(\widetilde{\boldsymbol{x}}_t\widetilde{\boldsymbol{x}}_t^{\top})\right)K_h\left(\tau_t-\tau\right)\right\| + \left(\frac{1}{T}\sup_{\tau,\tau_t}|K_h(\tau_t-\tau)|\right)\cdot\sum_{t=1}^{T}\left\|\boldsymbol{x}_t\boldsymbol{x}_t^{\top} - \widetilde{\boldsymbol{x}}_t\widetilde{\boldsymbol{x}}_t^{\top}\right\|$$

$$= O_P\left(\sqrt{\frac{\log T}{Th}}\right) + O_P\left(\frac{1}{Th}\right) = O_P\left(\sqrt{\frac{\log T}{Th}}\right).$$

Hence, in the following we will directly apply Lemma B.7 to time-varying VAR process.

For notational simplicity, let

$$\boldsymbol{S}_{T,k}(\tau) = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}^{\top}\left(\frac{\tau_t-\tau}{h}\right)^k K_h(\tau_t-\tau) \text{ for } 0 \leq k \leq 2,$$

$$\boldsymbol{M}(\tau_t) = \boldsymbol{A}(\tau_t) - \boldsymbol{A}(\tau) - \boldsymbol{A}^{(1)}(\tau)(\tau_t-\tau) - \frac{1}{2}\boldsymbol{A}^{(2)}(\tau)(\tau_t-\tau)^2.$$

We now begin our investigation. Since

$$\boldsymbol{x}_t = \boldsymbol{Z}_{t-1}^\top \mathrm{vec}\left(\boldsymbol{A}(\tau) + \boldsymbol{A}^{(1)}(\tau)(\tau_t - \tau) + \frac{1}{2}\boldsymbol{A}^{(2)}(\tau)(\tau_t - \tau)^2 + \boldsymbol{M}(\tau_t)\right) + \boldsymbol{\eta}_t,$$

we write

$$\mathrm{vec}(\widehat{\boldsymbol{A}}(\tau) - \boldsymbol{A}(\tau))$$

$$= \left(\frac{1}{T}\sum_{t=1}^T \boldsymbol{Z}_{t-1}\boldsymbol{Z}_{t-1}^\top K_h(\tau_t - \tau)\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^T \boldsymbol{Z}_{t-1}\boldsymbol{x}_t K_h(\tau_t - \tau)\right) - \mathrm{vec}\left(\boldsymbol{A}(\tau)\right)$$

$$= \left(\boldsymbol{S}_{T,0}^{-1}(\tau) \otimes \boldsymbol{I}_d\right)\left\{\left(\boldsymbol{S}_{T,1}(\tau)\otimes\boldsymbol{I}_d\right)h\mathrm{vec}\left(\boldsymbol{A}^{(1)}(\tau)\right) + \left(\boldsymbol{S}_{T,2}(\tau)\otimes\boldsymbol{I}_d\right)\frac{1}{2}h^2\mathrm{vec}\left(\boldsymbol{A}^{(2)}(\tau)\right)\right\}$$

$$+ \left(\boldsymbol{S}_{T,0}^{-1}(\tau)\otimes\boldsymbol{I}_d\right)\left(\frac{1}{T}\sum_{t=1}^T(\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}^\top\otimes\boldsymbol{I}_d)\mathrm{vec}\left(\boldsymbol{M}(\tau_t)\right)K_h(\tau_t - \tau)\right)$$

$$+ \left(\boldsymbol{S}_{T,0}^{-1}(\tau)\otimes\boldsymbol{I}_d\right)\left(\frac{1}{T}\sum_{t=1}^T(\boldsymbol{z}_{t-1}\otimes\boldsymbol{I}_d)\boldsymbol{\eta}_t K_h(\tau_t - \tau)\right) := \boldsymbol{I}_{T,1} + \boldsymbol{I}_{T,2} + \boldsymbol{I}_{T,3}.$$

By standard arguments for local constant estimator and the uniform convergence results in Lemma B.7, we have

$$\|\boldsymbol{I}_{T,1} + \boldsymbol{I}_{T,2}\| = O(h^2) + O_P(h\sqrt{\log T/(Th)})$$

uniformly over $\tau \in [h, 1-h]$. By Lemma B.8, we have $\boldsymbol{I}_{T,3} = O_P((\frac{\log T}{Th})^{\frac{1}{2}})$ uniformly over $\tau \in [0,1]$. Thus, the first result follows.

(2). We begin our investigation on the asymptotic normality by writing

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

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

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

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

$$:= \boldsymbol{I}_{T,4} + \boldsymbol{I}_{T,5} + \boldsymbol{I}_{T,6} + \boldsymbol{I}_{T,7}.$$

Let $c_T = h^2 + \sqrt{\frac{\log T}{Th}}$. By the first result, for $\forall \tau \in (0,1)$ we have

$$\left\|\frac{1}{Th}\sum_{t=1}^T(\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)(\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)^\top K\left(\frac{\tau_t - \tau}{h}\right)\right\|$$

$$\leq \sup_{\tau_t\in[h,1-h]}\|\widehat{\boldsymbol{A}}(\tau_t) - \boldsymbol{A}(\tau_t)\|^2 \cdot \frac{1}{Th}\sum_{t=1}^T \|\boldsymbol{z}_{t-1}\|^2 K\left(\frac{\tau_t - \tau}{h}\right) = O_P(c_T^2).$$

By Lemma B.8, $\boldsymbol{I}_{T,6}$ and $\boldsymbol{I}_{T,7}$ are both $o_P((Th)^{-1/2})$. Hence,

$$\sqrt{Th}\left(\frac{1}{Th}\sum_{t=1}^{T}\widehat{\boldsymbol{\eta}}_t\widehat{\boldsymbol{\eta}}_t^\top K\left(\frac{\tau_t-\tau}{h}\right)-\frac{1}{Th}\sum_{t=1}^{T}\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top K\left(\frac{\tau_t-\tau}{h}\right)-o_P(h^4)\right)=o_P(1).$$

The above development yields that

$$\sqrt{Th}\begin{pmatrix}\mathrm{vec}\left(\widehat{\boldsymbol{A}}(\tau)-\boldsymbol{A}(\tau)-\frac{1}{2}h^2\tilde{c}_2\boldsymbol{A}^{(2)}(\tau)\right)-o_P(h^2)\\[2mm]\mathrm{vech}\left(\widehat{\boldsymbol{\Omega}}(\tau)-\boldsymbol{\Omega}(\tau)-\frac{1}{2}h^2\tilde{c}_2\boldsymbol{\Omega}^{(2)}(\tau)\right)-o_P(h^2)\end{pmatrix}$$

$$=\begin{pmatrix}\left(\boldsymbol{\Sigma}^{-1}(\tau)\otimes\boldsymbol{I}_d\right)\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\boldsymbol{Z}_{t-1}\boldsymbol{\eta}_t K\left(\frac{\tau_t-\tau}{h}\right)\right)\\[2mm]\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\mathrm{vech}\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top-\boldsymbol{\Omega}(\tau_t)\right)K\left(\frac{\tau_t-\tau}{h}\right)\end{pmatrix}+o_P(1)$$

$$:=\boldsymbol{I}_{T,8}+o_P(1).$$

Below, we focus on $\boldsymbol{I}_{T,8}$. First, show $\mathrm{Var}(\boldsymbol{I}_{T,8})\to\boldsymbol{V}(\tau)$. Let

$$\mathrm{Var}(\boldsymbol{I}_{T,8})=\begin{pmatrix}\widetilde{\boldsymbol{V}}_{1,1}(\tau)&\widetilde{\boldsymbol{V}}_{2,1}^\top(\tau)\\[2mm]\widetilde{\boldsymbol{V}}_{2,1}(\tau)&\widetilde{\boldsymbol{V}}_{2,2}(\tau)\end{pmatrix},$$

where the definition of each block should be obvious. Moreover, simple algebra shows that $\widetilde{\boldsymbol{V}}_{i,j}(\tau)\to\boldsymbol{V}_{i,j}(\tau)$ for $i,j\in\{1,2\}$.

By construction, $\boldsymbol{I}_{T,8}$ is a summation of m.d.s., we thus use Lemma B.1 and Cramér-Wold device to prove its asymptotic normality. It suffices to show that $\boldsymbol{d}^\top\boldsymbol{I}_{T,8}\to_D N\left(\boldsymbol{0},\boldsymbol{d}^\top\boldsymbol{V}(\tau)\boldsymbol{d}\right)$ for any conformable unit vector $\boldsymbol{d}$. Let

$$\boldsymbol{Z}_{T,t}(\tau)=\frac{1}{\sqrt{Th}}\boldsymbol{d}^\top\begin{pmatrix}\left(\boldsymbol{\Sigma}^{-1}(\tau)\otimes\boldsymbol{I}_d\right)\left(\boldsymbol{Z}_{t-1}\boldsymbol{\eta}_t K\left(\frac{\tau_t-\tau}{h}\right)\right)\\[2mm]\mathrm{vech}\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top-\boldsymbol{\Omega}(\tau_t)\right)K\left(\frac{\tau_t-\tau}{h}\right)\end{pmatrix}.$$

By the law of large numbers of m.d.s., we have $\sum_{t=1}^{T}\boldsymbol{Z}_{T,t}^2(\tau)-\sum_{t=1}^{T}E(\boldsymbol{Z}_{T,t}^2(\tau)|\mathcal{F}_{t-1})\to_P 0$. Since conditional on $\mathcal{F}_{t-1}$ the third and fourth moments of $\boldsymbol{\epsilon}_t$ equal to the corresponding unconditional moments a.s., we can prove that $\sum_{t=1}^{T}E(\boldsymbol{Z}_{T,t}^2(\tau)|\mathcal{F}_{t-1})\to_P\boldsymbol{d}^\top\boldsymbol{V}(\tau)\boldsymbol{d}$.

Furthermore, for any $\nu>0$ and $\tau\in(0,1)$, similar to the proof of Theorem 2.2, we have

$$\sum_{t=1}^{T}E\left(\left(\boldsymbol{d}^\top\boldsymbol{Z}_{T,t}(\tau)\right)^2 I\left(|\boldsymbol{d}^\top\boldsymbol{Z}_{T,t}(\tau)|>\nu\right)\right)\to 0.$$

The result follows by Lemma B.1.

(3). By Lemma B.7 and the second result of Theorem 3.1, we have $\widehat{\boldsymbol{V}}_{1,1}(\tau)\to_P\boldsymbol{V}_{1,1}(\tau)$. Similar to the proof of the second result of this theorem, by the uniform convergence results of $\widehat{\boldsymbol{A}}(\tau)$, we can replace $\widehat{\boldsymbol{\eta}}_t$ with $\boldsymbol{\eta}_t$ in the following calculations. Therefore,

$$\widehat{\boldsymbol{V}}_{2,1}(\tau)=\frac{1}{Th}\sum_{t=1}^{T}\mathrm{vech}\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top\right)\boldsymbol{\eta}_t^\top\boldsymbol{Z}_{t-1}^\top K^2\left(\frac{\tau_t-\tau}{h}\right)\left(\boldsymbol{\Sigma}^{-1}(\tau)\otimes\boldsymbol{I}_d\right)+o_P(1)\to_P\boldsymbol{V}_{2,1}(\tau),$$

and

$$\widehat{\boldsymbol{V}}_{2,2}(\tau) = \frac{1}{Th} \sum_{t=1}^{T} \text{vech}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top) \text{vech}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top)^\top K^2 \left(\frac{\tau_t - \tau}{h} \right)$$

$$- \widetilde{v}_0 \text{vech} \left(\boldsymbol{\Omega}(\tau) \right) \text{vech} \left(\boldsymbol{\Omega}(\tau) \right)^\top + o_P(1) \to_P \boldsymbol{V}_{2,2}(\tau).$$

The proof is now completed. □

Proof of Theorem 3.2.

We prove that $\lim_{T \to \infty} \Pr \left(\text{IC}(\mathsf{p}) < \text{IC}(p) \right) = 0$ for all $\mathsf{p} \neq p$ and $\mathsf{p} \leq \mathsf{P}$.

Note that

$$\text{IC}(\mathsf{p}) - \text{IC}(p) = \log[\text{RSS}(\mathsf{p})/\text{RSS}(p)] + (\mathsf{p} - p)\chi_T.$$

For $\mathsf{p} < p$, Lemma B.9 implies that $\text{RSS}(\mathsf{p})/\text{RSS}(p) > 1 + \nu$ for some $\nu > 0$ with large probability for all large T. Thus, $\log[\text{RSS}(\mathsf{p})/\text{RSS}(p)] \geq \nu/2$ for large T. Because $\chi_T \to 0$, we have $\text{IC}(\mathsf{p}) - \text{IC}(p) \geq \nu/2 - (p - \mathsf{p})\chi_T \geq \nu/3$ for large T with large probability. Thus $\Pr \left(\text{IC}(\mathsf{p}) < \text{IC}(p) \right) \to 0$ for $\mathsf{p} < p$.

Next, consider $\mathsf{p} > p$. Lemma B.9 implies that $\log[\text{RSS}(\mathsf{p})/\text{RSS}(p)] = 1 + O_P(c_T \phi_T)$. Hence, $\log[\text{RSS}(\mathsf{p})/\text{RSS}(p)] = O_P(c_T \phi_T)$. Because $(\mathsf{p} - p)\chi_T \geq \chi_T$, which converges to zero at a slower rate than $c_T \phi_T$, it follows that

$$\Pr \left(\text{IC}(\mathsf{p}) < \text{IC}(p) \right) \leq \Pr \left(O_P(c_T \phi_T) + \chi_T < 0 \right) \to 0.$$

The proof is now completed. □

Proof of Theorem 3.3.

Given the joint distribution of $\text{vec}(\widehat{\boldsymbol{A}}(\tau))$ and $\text{vech}(\widehat{\boldsymbol{\Omega}}(\tau))$ in Theorem 3.1, Theorem 3.3 can by easily obtained by the Delta method since

$$\sqrt{Th} \text{vec} \left(\widehat{\boldsymbol{B}}_j(\tau) - \boldsymbol{B}_j(\tau) \right) = \sqrt{Th} \text{vec} \left(\boldsymbol{J} \widehat{\boldsymbol{\Phi}}^j(\tau) \boldsymbol{J}^\top \widehat{\boldsymbol{\omega}}(\tau) - \boldsymbol{J} \boldsymbol{\Phi}^j(\tau) \boldsymbol{J}^\top \boldsymbol{\omega}(\tau) \right).$$

Then, by standard arguments of the Delta method (see Lütkepohl, 2005, p.111), we can show that

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

where $\boldsymbol{\Sigma}_{\boldsymbol{B}_j}(\tau)$ has been defined in the body of the theorem. □

Online Supplementary Appendix B to "A Class of Time–Varying VMA(∞) Models"

This file includes the simulations, preliminary lemmas and proofs which are omitted in the main text. Specifically, the simulations are summarized in Appendix B.1; Appendix B.2 presents the preliminary lemmas which are helpful to derive the main results of the paper; Appendix B.3 includes the omitted proofs of the main results; the proofs of the secondary lemmas are presented in Appendix B.4.

Appendix B

B.1 Simulation

In this section, we exam the above theoretical findings using intensive simulation. The Epanechnikov kernel $K(u) = 0.75(1-u^2)I(|u| \leq 1)$ is adopted throughout the numerical studies of this paper for simplicity. For each estimation conducted below, we always select the number of lag by (3.7) by searching the estimate of p over a sufficiently large range, say $\{1, \ldots, \lfloor \sqrt{Th} \rfloor\}$. Moreover, for each given p, the bandwidth is selected by minimizing the following cross–validation criterion function

$$\text{CV}(h) = \sum_{t=1}^{T} \left(\boldsymbol{x}_t - \boldsymbol{a}_{-t}(\tau_t) - \sum_{j=1}^{\mathsf{p}} \boldsymbol{A}_{j,-t}(\tau_t) \boldsymbol{x}_{t-j} \right)^{\top} \left(\boldsymbol{x}_t - \boldsymbol{a}_{-t}(\tau_t) - \sum_{j=1}^{\mathsf{p}} \boldsymbol{A}_{j,-t}(\tau_t) \boldsymbol{x}_{t-j} \right), \qquad \text{(B.1)}$$

where $\boldsymbol{a}_{-t}(\cdot)$, and $\boldsymbol{A}_{j,-t}(\cdot)$ are obtained using (3.3) but leaving the tth observation out. Once $\widehat{\mathsf{p}}$ is obtained, the rest calculation is relatively straightforward.

We now start describing the data generating process. Let $\boldsymbol{\epsilon}_t$ be i.i.d. draws from $N(\boldsymbol{0}_{2\times 1}, \boldsymbol{I}_2)$ with $E\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{\top}\right) = \boldsymbol{I}_2$. Consider

$$\boldsymbol{x}_t = \boldsymbol{a}(\tau_t) + \boldsymbol{A}_1(\tau_t) \boldsymbol{x}_{t-1} + \boldsymbol{A}_2(\tau_t) \boldsymbol{x}_{t-2} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau_t) \boldsymbol{\epsilon}_t \quad \text{with} \quad t = 1, \ldots, T,$$

where

$$\boldsymbol{a}(\tau) = [0.5\sin(2\pi\tau), 0.5\cos(2\pi\tau)]^{\top},$$

$$\boldsymbol{A}_1(\tau) = \begin{pmatrix} 0.8\exp\left(-0.5+\tau\right) & 0.8(\tau-0.5)^3 \\ 0.8(\tau-0.5)^3 & 0.8+0.3\sin(\pi\tau) \end{pmatrix},$$

$$\boldsymbol{A}_2(\tau) = \begin{pmatrix} -0.2\exp\left(-0.5+\tau\right) & 0.8(\tau-0.5)^2 \\ 0.8(\tau-0.5)^2 & -0.4+0.3\cos(\pi\tau) \end{pmatrix},$$

$$\boldsymbol{\omega}(\tau) = \begin{pmatrix} 1.5+0.2\exp\left(0.5-\tau\right) & 0 \\ 0.2(1.5+0.5(\tau-0.5)^2)(1.5+0.2\exp\left(0.5-\tau\right)) & 1.5+0.5(\tau-0.5)^2 \end{pmatrix}.$$

We consider the sample size $T \in \{200, 400, 800\}$, and conduct 1000 replications for each choice of T.

1

Based on 1000 replications, we first report the percentages of $\widehat{\mathsf{p}} < 2$, $\widehat{\mathsf{p}} = 2$, and $\widehat{\mathsf{p}} > 2$ respectively in Table B.1 . It can be seen that the information criterion (3.7) performs reasonably well. The percentage associated to $\widehat{\mathsf{p}} = 2$ increases as the sample size goes up.

Table B.1: The percentages of $\widehat{\mathsf{p}} < 2$, $\widehat{\mathsf{p}} = 2$, and $\widehat{\mathsf{p}} > 2$

T	$\widehat{\mathsf{p}} < 2$	$\widehat{\mathsf{p}} = 2$	$\widehat{\mathsf{p}} > 2$
200	0.061	0.895	0.044
400	0.011	0.964	0.025
800	0.001	0.983	0.016

Next, we evaluate the estimates of $\boldsymbol{A}(\tau)$ and $\boldsymbol{\Omega}(\tau)$, and calculate the root mean square error (RMSE) as follows.

$$\left\{ \frac{1}{nT} \sum_{n=1}^{1000} \sum_{t=1}^{T} \|\widehat{\boldsymbol{\theta}}^{(n)}(\tau_t) - \boldsymbol{\theta}(\tau_t)\|^2 \right\}^{1/2}$$

for $\boldsymbol{\theta}(\cdot) \in \{\boldsymbol{A}(\cdot), \boldsymbol{\Omega}(\cdot)\}$, where $\widehat{\boldsymbol{\theta}}^{(n)}(\tau)$ is the estimate of $\boldsymbol{\theta}(\tau)$ for the n-th replication. Of interest, we also examine the finite sample coverage probabilities of the confidence intervals based on our asymptotic theories. In the following, we compute the average of coverage probabilities for grid points in $\{\tau_t, t = 1, \ldots, T\}$. The RMSEs and empirical coverage probabilities are reported in Table B.2. As shown in Table B.2, the RMSE decreases as the sample size goes up. The finite sample coverage probabilities are smaller than their nominal level (95%) for small T, but are fairly close to 95% as T increases.

Table B.2: The RMSEs and the empirical coverage probabilities with 95% nominal level.

T	$\boldsymbol{A}(\tau)$		$\boldsymbol{\Omega}(\tau)$	
	RMSE	Coverage rate	RMSE	Coverage rate
200	0.4937	0.9265	0.8228	0.8687
400	0.3703	0.9291	0.7143	0.9059
800	0.2785	0.9319	0.6205	0.9226

B.2   Preliminary Lemmas

We present the preliminary lemmas below, which help facilitate the development of the main results.

Lemma B.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 B.1 can be found in Hall and Heyde (1980).

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

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

Lemma B.2 is the Proposition 2.1 of Freedman (1975).

Lemma B.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 B.4. *Let Assumptions 1 and 2 hold. Suppose* $\{\boldsymbol{W}_{T,t}\}_{t=1}^{T}$ *is a sequence* $m \times d$ *deterministic matrices satisfying (1).* $\sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}\| = O(1)$, *(2).* $\max_{t\geq 1} \|\boldsymbol{W}_{T,t}\| = o(1)$, *and (3).* $\sum_{t=1}^{T-1} \|\boldsymbol{W}_{T,t+1} - \boldsymbol{W}_{T,t}\| = o(1)$. *As* $T \to \infty$,

$$\sum_{t=1}^{T} \boldsymbol{W}_{T,t} \left(\boldsymbol{x}_t - E(\boldsymbol{x}_t)\right) \to_P \boldsymbol{0} \quad and \quad \sum_{t=1}^{T} \boldsymbol{W}_{T,t} \left(\boldsymbol{x}_t \boldsymbol{x}_{t+p}^{\top} - E\left(\boldsymbol{x}_t \boldsymbol{x}_{t+p}^{\top}\right)\right) \to_P \boldsymbol{0},$$

where $m \geq 1$ *is fixed, and* p *is a fixed non–negative integer.*

Lemma B.5. *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 B.6. *Let the conditions of Lemma B.5 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* $\sup_{\tau \in [a,b]} \sum_{t=1}^{T-1} \|\boldsymbol{W}_{T,t+1}(\tau) - \boldsymbol{W}_{T,t}(\tau)\| = o(1)$. *As* $T \to \infty$

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

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

3

Lemma B.7. *Let Assumptions 2, 3, and 4 hold. As $T \to \infty$,*

1. *for $\tau \in (0,1)$,*

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_t \left(\frac{\tau_t - \tau}{h}\right)^k K_h(\tau_t - \tau) - \widetilde{c}_k \boldsymbol{\mu}(\tau) \to_P \boldsymbol{0},$$

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_{t+p}^{\top} \left(\frac{\tau_t - \tau}{h}\right)^k K_h(\tau_t - \tau) - \widetilde{c}_k \boldsymbol{\Sigma}_p(\tau) \to_P \boldsymbol{0},$$

where $\boldsymbol{\Sigma}_p(\tau) = \boldsymbol{\mu}(\tau)\boldsymbol{\mu}^{\top}(\tau) + \sum_{j=0}^{\infty} \boldsymbol{B}_j(\tau)\boldsymbol{B}_{j+p}^{\top}(\tau)$ for fixed integer $k, p \geq 0$;

2. *given $\frac{T^{1-\frac{2}{\delta}}h}{\log T} \to \infty$,*

$$\sup_{\tau \in [h, 1-h]} \left\| \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_t \left(\frac{\tau_t - \tau}{h}\right)^k K_h(\tau_t - \tau) - \widetilde{c}_k \boldsymbol{\mu}(\tau) \right\| = O_P\left(h^2 + \left(\frac{\log T}{Th}\right)^{\frac{1}{2}}\right);$$

3. *given $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$ and $\max_{t \geq 1} E[\|\boldsymbol{\epsilon}_t\|^4 | \mathcal{F}_{t-1}] < \infty$ a.s.,*

$$\sup_{\tau \in [h, 1-h]} \left\| \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_{t+p}^{\top} \left(\frac{\tau_t - \tau}{h}\right)^k K_h(\tau_t - \tau) - \widetilde{c}_k \boldsymbol{\Sigma}_p(\tau) \right\| = O_P\left(h^2 + \left(\frac{\log T}{Th}\right)^{\frac{1}{2}}\right).$$

Lemma B.8. *Let Assumptions 2, 4 and 5 hold. Suppose $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$ and $\max_{t \geq 1} E[\|\boldsymbol{\epsilon}_t\|^4 | \mathcal{F}_{t-1}] < \infty$ a.s. As $T \to \infty$,*

1. $\sup_{\tau \in [0,1]} \left\| \frac{1}{Th}\sum_{t=1}^{T} \boldsymbol{Z}_{t-1}\boldsymbol{\eta}_t K\left(\frac{\tau_t - \tau}{h}\right) \right\| = O_P\left(\left(\frac{\log T}{Th}\right)^{\frac{1}{2}}\right);$

2. $\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} \boldsymbol{\eta}_t (\boldsymbol{\eta}_t - \widehat{\boldsymbol{\eta}}_t)^{\top} K\left(\frac{\tau_t - \tau}{h}\right) = o_P(1)$ *for $\forall \tau \in (0,1)$.*

Lemma B.9. *Let Assumptions 2, 4 and 5 hold. Suppose $\frac{T^{1-\frac{4}{\delta}}h}{\log T} \to \infty$ and $\max_{t \geq 1} E[\|\boldsymbol{\epsilon}_t\|^4 | \mathcal{F}_{t-1}] < \infty$ a.s. As $T \to \infty$,*

1. *if $\mathsf{p} \geq p$, then $RSS(\mathsf{p}) = \frac{1}{T}\sum_{t=1}^{T} E\left(\boldsymbol{\eta}_t^{\top}\boldsymbol{\eta}_t\right) + O_P\left(c_T \phi_T\right)$;*

2. *if $\mathsf{p} < p$, then $RSS(\mathsf{p}) = \frac{1}{T}\sum_{t=1}^{T} E\left(\boldsymbol{\eta}_t^{\top}\boldsymbol{\eta}_t\right) + c + o_P(1)$ with some constant $c > 0$.*

B.3   Omitted Proofs of the Main Results

We present the omitted proofs of the main results in this section.

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 $\max_{t \geq 1} \|\prod_{i=0}^{j-1} \boldsymbol{\Phi}_{t-i}\| \leq M\rho^j$, which yields that

4

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

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

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

$$= \limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty}j\left\|\boldsymbol{J}\sum_{m=1}^{j}\left(\prod_{k=1}^{m-1}\boldsymbol{\Phi}_{t+2-k}\right)(\boldsymbol{\Phi}_{t+2-m} - \boldsymbol{\Phi}_{t+1-m})\left(\prod_{k=m}^{j}\boldsymbol{\Phi}_{t+1-k}\right)\boldsymbol{J}^{\top}\right\|$$

$$\leq M\sum_{j=1}^{\infty}j^{2}\rho^{j-1}\limsup_{T\to\infty}\sum_{t=1}^{T-1}\left\|\boldsymbol{\Phi}_{t+1} - \boldsymbol{\Phi}_{t}\right\| = O(1)$$

given the condition in Proposition 2.1.

Consider Example 2. Similar to Example 1,

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

In addition,

$$\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{b=1}^{\infty}b\left\|\boldsymbol{D}_{b,t+1} - \boldsymbol{D}_{b,t}\right\|$$

$$\leq \limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{b=1}^{\infty}b\sum_{j=\max(0,b-q)}^{b}\left\|\boldsymbol{B}_{j,t+1} - \boldsymbol{B}_{j,t}\right\|\left\|\boldsymbol{\Theta}_{b-j,t+1-j}\right\|$$

$$+ \limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{b=1}^{\infty}b\sum_{j=\max(0,b-q)}^{b}\left\|\boldsymbol{B}_{j,t}\right\|\left\|\boldsymbol{\Theta}_{b-j,t+1-j} - \boldsymbol{\Theta}_{b-j,t-j}\right\|$$

$$\leq M\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{b=1}^{\infty}b\left\|\boldsymbol{B}_{b,t+1} - \boldsymbol{B}_{b,t}\right\|$$

$$+ \sum_{b=1}^{\infty}b\sum_{j=\max(0,b-q)}^{b}\left\|\boldsymbol{B}_{j,t}\right\|\limsup_{T\to\infty}\sum_{t=1}^{T-1}\left\|\boldsymbol{\Theta}_{b-j,t+1-j} - \boldsymbol{\Theta}_{b-j,t-j}\right\| = O(1).$$

(2). By the condition of Proposition 2.1.3,

$$\max_{t\geq 1}\sum_{j=1}^{\infty}j\left\|\boldsymbol{B}_{j,t}\right\| \leq \max_{t\geq 1}\sum_{j=1}^{\infty}j\sum_{l=0}^{j}\left\|\boldsymbol{\Psi}_{l,t}\right\|\left\|\boldsymbol{\Theta}_{j-l,t-l}\right\|$$

$$= \max_{t\geq 1}\sum_{j=0}^{\infty}\sum_{l=j+1}^{\infty}l\left\|\boldsymbol{\Psi}_{j,t}\right\|\left\|\boldsymbol{\Theta}_{l-j,t-j}\right\| = \max_{t\geq 1}\sum_{j=0}^{\infty}\left\|\boldsymbol{\Psi}_{j,t}\right\|\sum_{l=1}^{\infty}(l+j)\left\|\boldsymbol{\Theta}_{l,t-j}\right\|$$

$$= \max_{t\geq 1}\sum_{j=0}^{\infty}j\left\|\boldsymbol{\Psi}_{j,t}\right\|\sum_{l=1}^{\infty}\left\|\boldsymbol{\Theta}_{l,t-j}\right\| + \max_{t\geq 1}\sum_{j=0}^{\infty}\left\|\boldsymbol{\Psi}_{j,t}\right\|\sum_{l=1}^{\infty}l\left\|\boldsymbol{\Theta}_{l,t-j}\right\| = O(1).$$

In addition,

$$\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty}j\left\|\boldsymbol{B}_{j,t+1}-\boldsymbol{B}_{j,t}\right\| \leq \limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty}j\sum_{l=0}^{j}\left\|\boldsymbol{\Psi}_{l,t+1}\right\|\left\|\boldsymbol{\Theta}_{j-l,t+1-l}-\boldsymbol{\Theta}_{j-l,t-l}\right\|$$
$$+\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty}j\sum_{l=0}^{j}\left\|\boldsymbol{\Psi}_{l,t+1}-\boldsymbol{\Psi}_{l,t}\right\|\left\|\boldsymbol{\Theta}_{j-l,t-l}\right\|$$
$$:= I_{T,1}+I_{T,2}.$$

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

$$I_{T,1} \leq \max_{t\geq 1}\sum_{j=0}^{\infty}j\left\|\boldsymbol{\Psi}_{j,t}\right\|\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{l=1}^{\infty}\left\|\boldsymbol{\Theta}_{l,t+1-j}-\boldsymbol{\Theta}_{l,t-j}\right\|$$
$$+\max_{t\geq 1}\sum_{j=0}^{\infty}\left\|\boldsymbol{\Psi}_{j,t}\right\|\limsup_{T\to\infty}\sum_{t=1}^{T-1}\sum_{l=1}^{\infty}l\left\|\boldsymbol{\Theta}_{l,t+1-j}-\boldsymbol{\Theta}_{l,t-j}\right\| = O(1).$$

The proof is now completed. □

Proof of Proposition 3.1.

Consider the VMA representation of $\boldsymbol{x}_t$:

$$\boldsymbol{x}_t = \boldsymbol{\mu}_t + \boldsymbol{B}_{0,t}\boldsymbol{\epsilon}_t + \boldsymbol{B}_{1,t}\boldsymbol{\epsilon}_{t-1} + \boldsymbol{B}_{2,t}\boldsymbol{\epsilon}_{t-2} + \cdots,$$

where $\boldsymbol{B}_{0,t}=\boldsymbol{\omega}(\tau_t)$, $\boldsymbol{B}_{j,t}=\boldsymbol{\Psi}_{j,t}\boldsymbol{\omega}(\tau_{t-j})$, $\boldsymbol{\Psi}_{j,t}=\boldsymbol{J}\prod_{m=0}^{j-1}\boldsymbol{\Phi}(\tau_{t-m})\boldsymbol{J}^{\top}$ for $j\geq 1$, $\boldsymbol{\mu}_t=\boldsymbol{a}(\tau_t)+\sum_{j=1}^{\infty}\boldsymbol{\Psi}_{j,t}\boldsymbol{a}(\tau_{t-j})$ and $\tau_{t-j}=\frac{t-j}{T}I(t\geq j)$.

First, we investigate the validity of VMA representation of $\boldsymbol{x}_t$ and $\widetilde{\boldsymbol{x}}_t$. Let ρ_A denote the largest eigenvalue of $\boldsymbol{\Phi}(\tau)$ uniformly over $\tau\in[0,1]$. Then, $\rho_A<1$ by Assumption 5.1. Similar to the proof of Proposition 2.4 in Dahlhaus and Polonik (2009), we have $\max_{t\geq 1}\left\|\prod_{m=0}^{j-1}\boldsymbol{\Phi}(\tau_{t-m})\right\| \leq M\rho_A^j$. It follows that $\|\mathrm{E}(\boldsymbol{x_t})\| \leq \sum_{j=0}^{\infty}\|\boldsymbol{\Psi}_{j,t}\|\cdot\|\boldsymbol{a}(\tau_{t-j})\| \leq M\sum_{j=0}^{\infty}\rho_A^j < \infty$ and

$$\|\mathrm{Var}(\boldsymbol{x_t})\| = \left\|\sum_{j=0}^{\infty}\boldsymbol{B}_{j,t}\boldsymbol{B}_{j,t}^{\top}\right\| \leq \sum_{j=0}^{\infty}\|\boldsymbol{B}_{j,t}\|^2 \leq M\sum_{j=0}^{\infty}\rho_A^{2j} < \infty.$$

Similarly, we have $\|\mathrm{E}(\widetilde{\boldsymbol{x}}_t)\|<\infty$ and $\|\mathrm{Var}(\widetilde{\boldsymbol{x}}_t)\|<\infty$.

Then, we need to verify that $\max_{t\geq 1}E\|\boldsymbol{x}_t-\widetilde{\boldsymbol{x}}_t\|=O(T^{-1})$. 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 have

$$\|\boldsymbol{B}_{j,t}-\boldsymbol{B}_j(\tau_t)\| = \left\|\boldsymbol{J}\prod_{m=0}^{j-1}\boldsymbol{\Phi}(\tau_{t-m})\boldsymbol{J}^{\top}\boldsymbol{\omega}(\tau_{t-j}) - \boldsymbol{J}\boldsymbol{\Phi}^j(\tau_t)\boldsymbol{J}^{\top}\boldsymbol{\omega}(\tau_t)\right\|$$

$$= \left\| \left(\boldsymbol{J} \prod_{m=0}^{j-1} \boldsymbol{\Phi}(\tau_{t-m}) \boldsymbol{J}^\top - \boldsymbol{J} \boldsymbol{\Phi}^j(\tau_t) \boldsymbol{J}^\top \right) \boldsymbol{\omega}(\tau_t) + \boldsymbol{J} \prod_{m=0}^{j-1} \boldsymbol{\Phi}(\tau_{t-m}) \boldsymbol{J}^\top (\boldsymbol{\omega}(\tau_{t-j}) - \boldsymbol{\omega}(\tau_t)) \right\|$$

$$\leq M \sum_{i=1}^{j-1} \left\| \boldsymbol{\Phi}^i(\tau_t)(\boldsymbol{\Phi}(\tau_{t-i}) - \boldsymbol{\Phi}(\tau_t)) \prod_{m=i+1}^{j-1} \boldsymbol{\Phi}(\tau_{t-m}) \right\| + M \rho_A^j \frac{j}{T}$$

$$\leq M \sum_{i=1}^{j-1} \frac{i}{T} \rho_A^{j-1} + M \rho_A^j \frac{j}{T} = O(T^{-1}),$$

which follows that

$$E \left\| \boldsymbol{x}_t - \widetilde{\boldsymbol{x}}_t \right\| \leq \sum_{j=1}^{\infty} \left\| \boldsymbol{\Psi}_{j,t} \boldsymbol{a}(\tau_{t-j}) - \boldsymbol{\Psi}_j(\tau_t) \boldsymbol{a}(\tau_t) \right\| + \sum_{j=1}^{\infty} \left\| \boldsymbol{B}_{j,t} - \boldsymbol{B}_j(\tau_t) \right\| \cdot E \left\| \boldsymbol{\epsilon}_t \right\|$$

$$\leq M \sum_{j=1}^{\infty} \left(\sum_{i=1}^{j-1} \frac{i}{T} \rho_A^{j-1} + \rho_A^j \frac{j}{T} \right) = O\left(T^{-1}\right).$$

Finally, we check whether the MA coefficients of $\widetilde{x}_t$ satisfy Assumption 3. For $\boldsymbol{\mu}(\tau)$, the series $\sum_{j=0}^{\infty} \boldsymbol{\Psi}_j(\tau) \boldsymbol{a}(\tau)$ converges uniformly on $[0,1] \in R$ since for every $\nu > 0$ there exists an $N_\nu > 0$ such that

$$\left\| \boldsymbol{\Psi}_{m+1}(\tau) \boldsymbol{a}(\tau) + \cdots + \boldsymbol{\Psi}_n(\tau) \boldsymbol{a}(\tau) \right\| \leq M \sum_{j=m+1}^{n} \rho_A^j < \nu$$

whenever $n > m > N_\nu$.

By the term-by-term differentiability theorem, we have $\boldsymbol{\mu}^{(1)}(\tau) = \sum_{j=0}^{\infty} \left(\boldsymbol{\Psi}_j^{(1)}(\tau) \boldsymbol{a}(\tau) + \boldsymbol{\Psi}_j(\tau) \boldsymbol{a}^{(1)}(\tau) \right)$, where $\boldsymbol{\Psi}_j^{(1)}(\tau) = \boldsymbol{J} \left(\sum_{i=0}^{j-1} \boldsymbol{\Phi}^i(\tau) \boldsymbol{\Phi}^{(1)}(\tau) \boldsymbol{\Phi}^{j-1-i}(\tau) \right) \boldsymbol{J}^\top$.

Therefore, we can conclude that $\boldsymbol{\mu}(\cdot)$ and $\boldsymbol{\Psi}_j(\cdot)$ is first-order continuously differentiable. Similarly, we can show the second-order continuously differentiability of $\boldsymbol{\mu}(\cdot)$ and $\boldsymbol{\Psi}_j(\cdot)$.

In addition, since $\sup_{\tau \in [0,1]} \left\| \boldsymbol{\Phi}^j(\tau) \right\| \leq M \rho_A^j$, we have

$$\sum_{j=0}^{\infty} j \left\| \boldsymbol{B}_j(\tau) \right\| = \sum_{j=0}^{\infty} j \left\| \boldsymbol{\Psi}_j(\tau) \boldsymbol{\omega}(\tau) \right\| \leq M \sum_{j=0}^{\infty} j \rho_A^j < \infty,$$

and

$$\sum_{j=0}^{\infty} j \left\| \boldsymbol{B}_j^{(1)}(\tau) \right\| = \sum_{j=0}^{\infty} j \left\| \boldsymbol{\Psi}_j^{(1)}(\tau) \boldsymbol{\omega}(\tau) + \boldsymbol{\Psi}_j(\tau) \boldsymbol{\omega}^{(1)}(\tau) \right\| \leq M \sum_{j=0}^{\infty} (j^2 \rho_A^{j-1} + j \rho_A^j) < \infty.$$

The proof is therefore completed.

$\square$

Proof of Theorem A.1.

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 B.4, we can show that $\|\boldsymbol{I}_{T,12}\| = O_P((T\widetilde{h})^{-1/2})$, which in connection of Assumption A.1 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

$$\boldsymbol{I}_{T,11} = \int_{-1}^{1} K(u) \int_{-1}^{1} K(v) \left(\boldsymbol{\mu}(\tau + v\widetilde{h} + uh) - \boldsymbol{\mu}(\tau + v\widetilde{h}) \right) dv du + O\left(\frac{1}{Th}\right)$$

$$= \frac{1}{2} h^2 \widetilde{c}_2 \boldsymbol{\mu}^{(2)}(\tau) + O(h^2(h + \widetilde{h})) + O\left(\frac{1}{Th}\right).$$

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

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

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

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

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

$$\frac{1}{\sqrt{Th}} \sum_{t=1}^{T} K\left(\frac{\tau_t - \tau}{h}\right) \boldsymbol{d}^\top \widehat{\boldsymbol{e}}_t \xi_t^* = Z_T^*(\tau) + \frac{1}{\sqrt{Th}} \sum_{t=1}^{T} K\left(\frac{\tau_t - \tau}{h}\right) \boldsymbol{d}^\top (\widehat{\boldsymbol{e}}_t - \boldsymbol{e}_t) \xi_t^*$$

$$= Z_T^*(\tau) + o_P(1),$$

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

$$EE^* \left\| \frac{1}{\sqrt{Th}} \sum_{t=1}^{T} K\left(\frac{\tau_t - \tau}{h}\right) \boldsymbol{d}^\top (\widehat{\boldsymbol{e}}_t - \boldsymbol{e}_t) \xi_t^* \right\|^2$$

$$\leq \max_{\lfloor T(\tau - h) \rfloor \leq t \leq \lceil T(\tau + h) \rceil} E \|\widehat{\boldsymbol{e}}_t - \boldsymbol{e}_t\|^2 \left(\frac{1}{Th} \sum_{t=1}^{T} \sum_{s=1}^{T} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_s - \tau}{h}\right) E^*(\xi_t^* \xi_s^*) \right)$$

$$= O\left(\widetilde{h}^4 + 1/(T\widetilde{h}) \right) O(l) = o(1),$$

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

In the following, we first show that

$$\text{Var}^*(Z_T^*(\tau))^2 = \tilde{v}_0 \boldsymbol{d}^\top \left\{ \sum_{j=0}^{\infty} \boldsymbol{B}_j(\tau) \right\} \left\{ \sum_{j=0}^{\infty} \boldsymbol{B}_j^\top(\tau) \right\} \boldsymbol{d} + o_P(1)$$

and then prove its normality by blocking techniques.

Condition on the original sample, we have

$$E^*(Z_T^*(\tau))^2$$
$$= \frac{1}{Th} \sum_{t=1}^{T} \sum_{s=1}^{T} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_s - \tau}{h}\right) \boldsymbol{d}^\top \boldsymbol{e}_t \boldsymbol{e}_s^\top \boldsymbol{d} E^*(\xi_t^* \xi_s^*)$$
$$= \frac{1}{Th} \sum_{t=1}^{T} K^2\left(\frac{\tau_t - \tau}{h}\right) \boldsymbol{d}^\top \boldsymbol{e}_t \boldsymbol{e}_t^\top \boldsymbol{d} + \frac{1}{Th} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) \boldsymbol{d}^\top \boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top \boldsymbol{d} a(i/l)$$
$$+ \frac{1}{Th} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) \boldsymbol{d}^\top \boldsymbol{e}_{t+i} \boldsymbol{e}_t^\top \boldsymbol{d} a(i/l). \tag{B.2}$$

For the first term on the right hand side of (B.2), by Lemma B.4, it is straightforward to obtain that

$$\frac{1}{Th} \sum_{t=1}^{T} K^2\left(\frac{\tau_t - \tau}{h}\right) \boldsymbol{d}^\top \boldsymbol{e}_t \boldsymbol{e}_t^\top \boldsymbol{d} = \frac{1}{Th} \sum_{t=1}^{T} K^2\left(\frac{\tau_t - \tau}{h}\right) \boldsymbol{d}^\top E\left(\boldsymbol{e}_t \boldsymbol{e}_t^\top\right) \boldsymbol{d} + o_P(1).$$

For the second and third terms on the right hand side of (B.2), we have

$$E \left\| \frac{1}{Th} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) \left(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top - E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top)\right) a(i/l) \right\|$$
$$\leq \frac{1}{Th} \sum_{i=1}^{T-1} a(i/l) E \left\| \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) \left(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top - E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top)\right) \right\|.$$

We now take a careful look at $E \left\| \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) \left(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top - E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top)\right) \right\|$. Simple algebra shows that $E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top) = \sum_{j=0}^{\infty} \boldsymbol{B}_{j,t} \boldsymbol{B}_{j+i,t+i}^\top$. Applying vector transformation, we can write

$$\frac{1}{Th} \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) \text{vec}\left[\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top - E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top)\right]$$
$$= \frac{1}{Th} \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) \sum_{j=0}^{\infty} (\boldsymbol{B}_{j+i,t+i} \otimes \boldsymbol{B}_{j,t}) \text{vec}\left[\boldsymbol{\epsilon}_{t-j} \boldsymbol{\epsilon}_{t-j}^\top - \boldsymbol{I}_d\right]$$
$$+ \frac{1}{Th} \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) \sum_{j=0}^{\infty} \sum_{m=0,\neq j+i}^{\infty} (\boldsymbol{B}_{m,t+i} \otimes \boldsymbol{B}_{j,t}) \text{vec}\left[\boldsymbol{\epsilon}_{t-j} \boldsymbol{\epsilon}_{t+i-m}^\top\right]$$
$$:= \boldsymbol{I}_{T,3} + \boldsymbol{I}_{T,4}.$$

Let $\boldsymbol{w}_t = \text{vec}\left[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top - \boldsymbol{I}_d\right]$. By Assumption 2, $E\|\boldsymbol{w}_t\|^{\delta/2} \leq 4E\|\boldsymbol{\epsilon}_t\|^\delta < \infty$ for some $\delta > 2$, which implies that $\{\boldsymbol{w}_t\}$ is uniformly integrable. Hence, for every $\nu > 0$, there exists a $\lambda_\nu > 0$ such that $E\|\boldsymbol{w}_t I(\|\boldsymbol{w}_t\| > \lambda_\nu)\| < \nu$. Define $\boldsymbol{w}_{1,t} = \boldsymbol{w}_t I(\|\boldsymbol{w}_t\| \leq \lambda_\nu)$ and $\boldsymbol{w}_{2,t} = \boldsymbol{w}_t - \boldsymbol{w}_{1,t} = \boldsymbol{w}_t I(\|\boldsymbol{w}_t\| > \lambda_\nu)$.

Similar to the proof of Theorem 2.22 in Hall and Heyde (1980),

$$
\begin{aligned}
E\|\boldsymbol{I}_{T,3}\| \leq\ & E\left\|\frac{1}{Th}\sum_{t=1}^{T-i} K\left(\frac{\tau_t-\tau}{h}\right) K\left(\frac{\tau_{t+i}-\tau}{h}\right)\right. \\
& \left.\cdot\sum_{j=0}^{\infty}\left(\boldsymbol{B}_{j+i,t+i}\otimes\boldsymbol{B}_{j,t}\right)\left(\boldsymbol{w}_{1,t-j}-E\left(\boldsymbol{w}_{1,t-j}|\mathcal{F}_{t-j-1}\right)\right)\right\| \\
& + E\left\|\frac{1}{Th}\sum_{t=1}^{T-i} K\left(\frac{\tau_t-\tau}{h}\right) K\left(\frac{\tau_{t+i}-\tau}{h}\right)\right. \\
& \left.\cdot\sum_{j=0}^{\infty}\left(\boldsymbol{B}_{j+i,t+i}\otimes\boldsymbol{B}_{j,t}\right)\left(\boldsymbol{w}_{2,t-j}-E\left(\boldsymbol{w}_{2,t-j}|\mathcal{F}_{t-j-1}\right)\right)\right\| \\
:=\ & I_{T,31}+I_{T,32}.
\end{aligned}
$$

For $I_{T,31}$,

$$
\begin{aligned}
I_{T,31} \leq\ & \frac{1}{Th}\sum_{j=0}^{\infty}\left\{ E\left\|\sum_{t=1}^{T-i} K\left(\frac{\tau_t-\tau}{h}\right) K\left(\frac{\tau_{t+i}-\tau}{h}\right)\right.\right. \\
& \left.\left.\cdot\left(\boldsymbol{B}_{j+i,t+i}\otimes\boldsymbol{B}_{j,t}\right)\left(\boldsymbol{w}_{1,t-j}-E\left(\boldsymbol{w}_{1,t-j}|\mathcal{F}_{t-j-1}\right)\right)\right\|^2\right\}^{\frac{1}{2}} \\
\leq\ & \frac{1}{Th}\sum_{j=0}^{\infty}\left\{\sum_{t=1}^{T-i} K^2\left(\frac{\tau_t-\tau}{h}\right) K^2\left(\frac{\tau_{t+i}-\tau}{h}\right)\right. \\
& \left.\cdot\|\boldsymbol{B}_{j+i,t+i}\|^2\|\boldsymbol{B}_{j,t}\|^2 E\|\boldsymbol{w}_{1,t-j}\|^2\right\}^{\frac{1}{2}} \\
\leq\ & \frac{1}{\sqrt{Th}}\max_{t\geq 1}\left(\sum_{j=0}^{\infty}\|\boldsymbol{B}_{j,t}\|\,\|\boldsymbol{B}_{j+i,t+i}\|\right) \\
& \cdot\left\{\frac{1}{Th}\sum_{t=1}^{T-i} K^2\left(\frac{\tau_t-\tau}{h}\right) K^2\left(\frac{\tau_{t+i}-\tau}{h}\right) E\|\boldsymbol{w}_{1,t-j}\|^2\right\}^{\frac{1}{2}} \\
:=\ & \beta_i\phi_{T,1}(\tau),
\end{aligned}
$$

where $\beta_i=\max_{t\geq 1}\sum_{j=0}^{\infty}\|\boldsymbol{B}_{j,t}\|\,\|\boldsymbol{B}_{j+i,t+i}\|$ satisfying

$$
\sum_{i=1}^{\infty}\beta_i \leq \sum_{i=1}^{\infty}\max_{t\geq 1}\sum_{j=0}^{\infty}\|\boldsymbol{B}_{j,t}\|\,\|\boldsymbol{B}_{j+i,t+i}\| \leq \left(\sup_{\tau}\sum_{j=0}^{\infty}\|\boldsymbol{B}_j(\tau)\|\right)^2 < \infty,
$$

and $\phi_{T,1}(\tau)=O\left(\frac{\lambda_\nu}{\sqrt{Th}}\right)$. For $I_{T,32}$,

$$
\begin{aligned}
I_{T,32} \leq\ & 2\max_{t\geq 1}\sum_{j=0}^{\infty}\|\boldsymbol{B}_{j,t}\|\,\|\boldsymbol{B}_{j+i,t+i}\|\frac{1}{Th}\sum_{t=1}^{T-i} K\left(\frac{\tau_t-\tau}{h}\right) K\left(\frac{\tau_{t+i}-\tau}{h}\right) E\|\boldsymbol{w}_{2,t-j}\| \\
:=\ & \beta_i\phi_{T,2}(\tau),
\end{aligned}
$$

where $\phi_{T,2}(\tau) = O(\nu)$. As we can make ν arbitrarily small, it follows that

$$\phi_T = \sup_{\tau \in [0,1]} (\phi_{T,1}(\tau) + \phi_{T,2}(\tau)) \to 0$$

as $T \to \infty$.

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

$$E\|\boldsymbol{I}_{T,4}\| \leq \frac{1}{\sqrt{Th}} \left(E \left\| \frac{1}{\sqrt{Th}} \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) \right. \right.$$

$$\left. \left. \cdot \sum_{j=0}^{\infty} \sum_{m=0,\neq j+i}^{\infty} (\boldsymbol{B}_{m,t+i} \otimes \boldsymbol{B}_{j,t}) \operatorname{vec}\left[\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t+i-m}^{\top}\right] \right\|^2 \right)^{\frac{1}{2}}$$

$$\leq O(1)\frac{1}{\sqrt{Th}},$$

since $E\left(\operatorname{vec}\left[\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t+i-m}^{\top}\right] \operatorname{vec}\left[\boldsymbol{\epsilon}_{s-j}\boldsymbol{\epsilon}_{s+i-m}^{\top}\right]^{\top} \right)$, $m \neq j+i$ can only be non-zero if $t = s$.

Based on the above development and $\sum_{i=0}^{\infty} a(i/l) = \sum_{i=0}^{l} a(i/l) = O(l)$, we conclude that

$$E \left\| \frac{1}{Th} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) \left(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^{\top} - E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^{\top})\right) a(i/l) \right\|$$

$$\leq M\phi_T \sum_{i=0}^{\infty} \beta_i + \frac{M}{\sqrt{Th}} \sum_{i=0}^{\infty} a(i/l) \leq \frac{Ml}{\sqrt{Th}} = o(1)$$

since $\sum_{i=1}^{\infty} \beta_i < \infty$ and $\lim_{T \to \infty} \phi_T = 0$.

We now just need to focus on $\frac{1}{Th} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^{\top}) a(i/l)$. Note that

$$\frac{1}{Th} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^{\top}) a(i/l)$$

$$= \frac{1}{Th} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^{\top})$$

$$+ \frac{1}{Th} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^{\top})(a(i/l) - 1).$$

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

$$\frac{1}{Th} \sum_{t=1}^{T} \sum_{s=1}^{T} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_s - \tau}{h}\right) E(\boldsymbol{e}_t \boldsymbol{e}_s^{\top})$$

$$= \operatorname{Var}\left(\frac{1}{\sqrt{Th}} \sum_{t=1}^{T} K\left(\frac{\tau_t - \tau}{h}\right) \boldsymbol{e}_t \right) \to \widetilde{v}_0 \left\{ \sum_{j=0}^{\infty} \boldsymbol{B}_j(\tau) \right\} \left\{ \sum_{j=0}^{\infty} \boldsymbol{B}_j^{\top}(\tau) \right\}$$

by the proof of Theorem 2.2.

Let d_T satisfy $\frac{1}{d_T} + \frac{d_T^2}{l} \to 0$. The second term is then bounded by

$$\left\| \frac{1}{Th} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right) E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top)(a(i/l) - 1) \right\|$$

$$\leq M \sum_{i=1}^{d_T} \max_t \left\| E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top) \right\| |a(i/l) - 1| + M \sum_{i=d_T+1}^{\infty} \max_t \left\| E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top) \right\| |a(i/l) - 1|$$

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

since

$$\sum_{i=1}^{d_T} (1 - a(i/l)) \leq \sum_{i=1}^{d_T} (-a^{(1)}(0)i/l + o(i/l)) \leq M d_T^2/l = o(1),$$

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

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

$$X_{T,j}^*(\tau) = \frac{1}{\sqrt{Th}} \sum_{t=B_j+1}^{B_j+r_1} K\left(\frac{\tau_t - \tau}{h}\right) \boldsymbol{d}^\top \boldsymbol{e}_t \xi_t^*, \quad Y_{T,j}^*(\tau) = \frac{1}{\sqrt{Th}} \sum_{t=B_j+r_1+1}^{B_j+r_1+r_2} K\left(\frac{\tau_t - \tau}{h}\right) \boldsymbol{d}^\top \boldsymbol{e}_t \xi_t^*,$$

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 $r_1/(Th) + l/(r_1) \to 0$ 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

$$E E^* \left(\sum_{j=1}^k Y_{T,j}^*(\tau) \right)^2 = E \left(\sum_{j=1}^k E^*(Y_{T,j}^*(\tau))^2 \right)$$

$$\leq \frac{1}{Th} \sum_{i=-r_2+1}^{r_2-1} a(i/l) \max_t \left\| E(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top) \right\| \sum_{j=1}^k \sum_{t=B_j+r_1+1}^{B_j+r_1+r_2-|i|} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+|i|} - \tau}{h}\right)$$

$$\leq M \frac{1}{Th} \max_{0 \leq i \leq r_2-1} \sum_{j=1}^k \sum_{t=B_j+r_1+1}^{B_j+r_1+r_2-i} K\left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_{t+i} - \tau}{h}\right)$$

$$\leq M \frac{k r_2 h}{Th} \leq M \frac{r_2}{r_1} = o(1).$$

We employ the 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 $\widetilde{v}_0 \boldsymbol{d}^\top \left\{ \sum_{j=0}^{\infty} \boldsymbol{B}_j(\tau) \right\} \left\{ \sum_{j=0}^{\infty} \boldsymbol{B}_j^\top(\tau) \right\} \boldsymbol{d}$. We then need to verify that for every $\nu > 0$,

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

Conditional on original sample, by Hölder's inequality, Chebyshev's inequality and Minkowski's inequality, we have

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

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

$$
= \nu^{\frac{2-\delta}{2}} \sum_{j=1}^{k} \frac{E^{*}(X_{T,j}^{*}(\tau))^{\delta}}{\left(E^{*}(\sum_{j=1}^{k} X_{T,j}^{*}(\tau))^{2}\right)^{\frac{\delta}{2}}} \leq \nu^{\frac{2-\delta}{2}} \sum_{j=1}^{k} \frac{\sum_{t=B_{j}+1}^{B_{j}+r_{1}}\left(\frac{1}{\sqrt{Th}} K\left(\frac{\tau_{t}-\tau}{h}\right) \boldsymbol{d}^{\top} \boldsymbol{e}_{t}\right)^{\delta} E^{*}(\xi_{t}^{*})^{\delta}}{\left(E^{*}(\sum_{j=1}^{k} X_{T,j}^{*}(\tau))^{2}\right)^{\frac{\delta}{2}}}
$$

$$
\leq \nu^{\frac{2-\delta}{2}} \frac{1}{(Th)^{\frac{\delta-2}{\delta}}} \frac{\frac{1}{Th} \sum_{t=1}^{T}\left(K\left(\frac{\tau_{t}-\tau}{h}\right) \boldsymbol{d}^{\top} \boldsymbol{e}_{t}\right)^{\delta} E^{*}(\xi_{t}^{*})^{\delta}}{\left(E^{*}(\sum_{j=1}^{k} X_{T,j}^{*}(\tau))^{2}\right)^{\frac{\delta}{2}}} = O_{P}\left(\frac{1}{(Th)^{\frac{\delta-2}{\delta}}}\right) = o_{P}(1).
$$

The proof is now completed. $\qquad\qquad\square$

B.4   Proofs of the Preliminary Lemmas

Proof of Lemma B.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

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

$$
= \sum_{j=0}^{\infty}\left(L^{j+1} \sum_{k=j+2}^{\infty}(\boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t}) - L^{j+1} \sum_{k=j+1}^{\infty}(\boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t})\right)
$$

$$
+ \sum_{k=1}^{\infty}(\boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t})
$$

$$
= -\sum_{j=0}^{\infty} L^{j+1}(\boldsymbol{B}_{j+1+r,t} \otimes \boldsymbol{B}_{j+1,t}) + \sum_{k=1}^{\infty}(\boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t})
$$

$$
= -\sum_{j=0}^{\infty} L^{j}(\boldsymbol{B}_{j+r,t} \otimes \boldsymbol{B}_{j,t}) + \sum_{k=0}^{\infty}(\boldsymbol{B}_{k+r,t} \otimes \boldsymbol{B}_{k,t})
$$

$$
= \mathbb{B}_{t}^{r}(1) - \mathbb{B}_{t}^{r}(L).
$$

(3). By Assumption 1,

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

13

(4). By Assumption 1,

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

$$= \sum_{t=1}^{T-1} \sum_{j=1}^{\infty} j \, \|\boldsymbol{B}_{j,t+1} - \boldsymbol{B}_{j,t}\| < \infty.$$

(5). By Assumption 1,

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

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

(6). Write

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

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

$$\le \max_{t \ge 1} \left(\sum_{r=1}^{\infty} \|\boldsymbol{B}_{r,t}\| \right) \cdot \left(\sum_{j=1}^{\infty} j \, \|\boldsymbol{B}_{j,t}\| \right) < \infty.$$

(7). Write

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

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

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

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

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

The proof is now completed. $\qquad \square$

Proof of Lemma B.4.

By Lemma B.3, we have

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

14

which yields that

$$\sum_{t=1}^{T} \boldsymbol{W}_{T,t} \left(\boldsymbol{x}_t - E(\boldsymbol{x}_t)\right) = \sum_{t=1}^{T} \boldsymbol{W}_{T,t}\mathbb{B}_t(1)\boldsymbol{\epsilon}_t + \boldsymbol{W}_{T,1}\widetilde{\mathbb{B}}_1(L)\boldsymbol{\epsilon}_0 - \boldsymbol{W}_{T,T}\widetilde{\mathbb{B}}_T(L)\boldsymbol{\epsilon}_T$$

$$+ \sum_{t=1}^{T-1} \left(\boldsymbol{W}_{T,t+1}\widetilde{\mathbb{B}}_{t+1}(L) - \boldsymbol{W}_{T,t}\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\epsilon}_t$$

$$:= \boldsymbol{I}_{T,1} + \boldsymbol{I}_{T,2} + \boldsymbol{I}_{T,3} + \boldsymbol{I}_{T,4}.$$

For $\boldsymbol{I}_{T,1}$,

$$E\left\|\sum_{t=1}^{T} \boldsymbol{W}_{T,t}\mathbb{B}_t(1)\boldsymbol{\epsilon}_t\right\|^2 = \mathrm{tr}\left(\sum_{t=1}^{T} \boldsymbol{W}_{T,t}\mathbb{B}_t(1)E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top)\mathbb{B}_t^\top(1)\boldsymbol{W}_{T,t}^\top\right)$$

$$\leq M\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}\|^2 \leq M\max_{t\geq 1}\|\boldsymbol{W}_{T,t}\|\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}\| = o(1).$$

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

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

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

$$\sum_{t=1}^{T-1} \left(\boldsymbol{W}_{T,t+1}\widetilde{\mathbb{B}}_{t+1}(L) - \boldsymbol{W}_{T,t}\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\epsilon}_t$$

$$= \sum_{t=1}^{T-1} \left(\boldsymbol{W}_{T,t+1} - \boldsymbol{W}_{T,t}\right)\widetilde{\mathbb{B}}_{t+1}(L)\boldsymbol{\epsilon}_t + \sum_{t=1}^{T-1} \boldsymbol{W}_{T,t}\left(\widetilde{\mathbb{B}}_{t+1}(L) - \widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\epsilon}_t. \tag{B.3}$$

Note that for the first term on the right hand side of (B.3)

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

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

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

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

$$\leq \max_{t\geq 1}E\|\boldsymbol{\epsilon}_t\| \cdot \max_{t\geq 1}\|\boldsymbol{W}_{T,t}\|\sum_{t=1}^{T-1}\|\widetilde{\mathbb{B}}_{t+1}(1) - \widetilde{\mathbb{B}}_t(1)\|$$

$$\leq M\max_{t\geq 1}\|\boldsymbol{W}_{T,t}\|\sum_{t=1}^{T-1}\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}\|\boldsymbol{B}_{j,t+1} - \boldsymbol{B}_{j,t}\|$$

$$= M\max_{t\geq 1}\|\boldsymbol{W}_{T,t}\|\sum_{t=1}^{T-1}\sum_{j=1}^{\infty}j\|\boldsymbol{B}_{j,t+1} - \boldsymbol{B}_{j,t}\| = o(1).$$

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

15

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

$$\boldsymbol{x}_t \boldsymbol{x}_t^{\top} = \boldsymbol{\mu}_t \boldsymbol{\mu}_t^{\top} + \boldsymbol{\mu}_t \sum_{j=0}^{\infty} \boldsymbol{\epsilon}_{t-j}^{\top} \boldsymbol{B}_{j,t}^{\top} + \sum_{j=0}^{\infty} \boldsymbol{B}_{j,t} \boldsymbol{\epsilon}_{t-j} \boldsymbol{\mu}_t^{\top} + \sum_{j=0}^{\infty} \boldsymbol{B}_{j,t} \boldsymbol{\epsilon}_{t-j} \boldsymbol{\epsilon}_{t-j}^{\top} \boldsymbol{B}_{j,t}^{\top}$$

$$+ \sum_{r=1}^{\infty} \sum_{j=0}^{\infty} \boldsymbol{B}_{j,t} \boldsymbol{\epsilon}_{t-j} \boldsymbol{\epsilon}_{t-j-r}^{\top} \boldsymbol{B}_{j+r,t}^{\top} + \sum_{r=1}^{\infty} \sum_{j=0}^{\infty} \boldsymbol{B}_{j+r,t} \boldsymbol{\epsilon}_{t-j-r} \boldsymbol{\epsilon}_{t-j}^{\top} \boldsymbol{B}_{j,t}^{\top},$$

which yields that

$$\mathrm{vec} \left[\boldsymbol{W}_{T,t} \left(\boldsymbol{x}_t \boldsymbol{x}_t^{\top} - E\left(\boldsymbol{x}_t \boldsymbol{x}_t^{\top} \right) \right) \right]$$

$$= (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \sum_{j=0}^{\infty} (\boldsymbol{B}_{j,t} \otimes \boldsymbol{\mu}_t) \, \boldsymbol{\epsilon}_{t-j} + (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \sum_{j=0}^{\infty} (\boldsymbol{\mu}_t \otimes \boldsymbol{B}_{j,t}) \, \boldsymbol{\epsilon}_{t-j}$$

$$+ (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \sum_{j=0}^{\infty} (\boldsymbol{B}_{j,t} \otimes \boldsymbol{B}_{j,t}) \, \mathrm{vec}[\boldsymbol{\epsilon}_{t-j} \boldsymbol{\epsilon}_{t-j}^{\top} - \boldsymbol{I}_d]$$

$$+ (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \sum_{r=1}^{\infty} \sum_{j=0}^{\infty} (\boldsymbol{B}_{j+r,t} \otimes \boldsymbol{B}_{j,t}) \mathrm{vec}[\boldsymbol{\epsilon}_{t-j} \boldsymbol{\epsilon}_{t-j-r}^{\top}]$$

$$+ (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \sum_{r=1}^{\infty} \sum_{j=0}^{\infty} (\boldsymbol{B}_{j,t} \otimes \boldsymbol{B}_{j+r,t}) \mathrm{vec}[\boldsymbol{\epsilon}_{t-j-r} \boldsymbol{\epsilon}_{t-j}^{\top}].$$

Consequently, we obtain

$$\left\| \sum_{t=1}^{T} \boldsymbol{W}_{T,t} \left(\boldsymbol{x}_t \boldsymbol{x}_t^{\top} - E\left(\boldsymbol{x}_t \boldsymbol{x}_t^{\top} \right) \right) \right\|$$

$$\leq 2 \left\| \sum_{t=1}^{T} (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \sum_{j=0}^{\infty} (\boldsymbol{\mu}_t \otimes \boldsymbol{B}_{j,t}) \, \boldsymbol{\epsilon}_{t-j} \right\|$$

$$+ \left\| \sum_{t=1}^{T} (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \sum_{j=0}^{\infty} (\boldsymbol{B}_{j,t} \otimes \boldsymbol{B}_{j,t}) \, \mathrm{vec}[\boldsymbol{\epsilon}_{t-j} \boldsymbol{\epsilon}_{t-j}^{\top} - \boldsymbol{I}_d] \right\|$$

$$+ 2 \left\| \sum_{t=1}^{T} (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \sum_{r=1}^{\infty} \sum_{j=0}^{\infty} (\boldsymbol{B}_{j+r,t} \otimes \boldsymbol{B}_{j,t}) \mathrm{vec}[\boldsymbol{\epsilon}_{t-j} \boldsymbol{\epsilon}_{t-j-r}^{\top}] \right\|$$

$$:= I_{T,5} + I_{T,6} + I_{T,7}.$$

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 $o_P(1)$.

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

$$I_{T,6} \leq \left\| \sum_{t=1}^{T} (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}) \, \widetilde{\mathbb{B}}_t^0(1) \left(\mathrm{vec}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{\top}) - \mathrm{vec}(\boldsymbol{I}_d) \right) \right\|$$

$$+ \left\| (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,1}) \, \widetilde{\mathbb{B}}_1^0(L) \mathrm{vec}(\boldsymbol{\epsilon}_0 \boldsymbol{\epsilon}_0^{\top}) \right\| + \left\| (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,T}) \, \widetilde{\mathbb{B}}_T^0(L) \mathrm{vec}(\boldsymbol{\epsilon}_T \boldsymbol{\epsilon}_T^{\top}) \right\|$$

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

$$:= I_{T,61} + I_{T,62} + I_{T,63} + I_{T,64}.$$

Let $\boldsymbol{Z}_t = \mathrm{vec}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{\top} - \boldsymbol{I}_d)$ for notational simplicity. By Assumption 2, for any $\nu > 0$, there exists $\lambda_\nu > 0$

that for all t, $E\left[\|\boldsymbol{Z}_t\| \cdot I\left(\|\boldsymbol{Z}_t\| > \lambda_\nu\right)\right] < \nu$. Then let further $\boldsymbol{Z}_{1,t} = \boldsymbol{Z}_t \cdot I\left(\|\boldsymbol{Z}_t\| \leq \lambda_\nu\right)$ and $\boldsymbol{Z}_{2,t} = \boldsymbol{Z}_t - \boldsymbol{Z}_{1,t}$. We are now ready to write

$$I_{T,61} \leq \left\|\sum_{t=1}^{T} \left(\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}\right) \widetilde{\mathbb{B}}_t^0(1)\left(\boldsymbol{Z}_{1,t} - E\left(\boldsymbol{Z}_{1,t}|\mathcal{F}_{t-1}\right)\right)\right\|$$
$$+ \left\|\sum_{t=1}^{T} \left(\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}\right) \widetilde{\mathbb{B}}_t^0(1)\left(\boldsymbol{Z}_{2,t} - E\left(\boldsymbol{Z}_{2,t}|\mathcal{F}_{t-1}\right)\right)\right\|. \tag{B.4}$$

For the first term on the right hand side of (B.4), by Lemma B.3, write

$$E\left\|\sum_{t=1}^{T} \left(\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}\right) \widetilde{\mathbb{B}}_t^0(1)\left(\boldsymbol{Z}_{1,t} - E\left(\boldsymbol{Z}_{1,t}|\mathcal{F}_{t-1}\right)\right)\right\|^2$$

$$= \text{tr}\left\{ \sum_{t=1}^{T}\sum_{s=1}^{T}(\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t})\widetilde{\mathbb{B}}_t^0(1)E\left(\left(\boldsymbol{Z}_{1,t} - E\left(\boldsymbol{Z}_{1,t}|\mathcal{F}_{t-1}\right)\right)\left(\boldsymbol{Z}_{1,s} - E\left(\boldsymbol{Z}_{1,s}|\mathcal{F}_{s-1}\right)\right)^{\top}\right) \right.$$

$$\left. \cdot\widetilde{\mathbb{B}}_s^{0,\top}(1)(\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,s}^{\top}) \right\}$$

$$\leq M\left(\max_{t\geq 1}\sum_{j=0}^{\infty}\left\|\widetilde{\boldsymbol{B}}_{j,t}^0\right\|\right)^2 \sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}\|^2\, E\|\boldsymbol{Z}_{1,t}\|^2 \leq M\lambda_\nu^2 \max_{t\geq 1}\|\boldsymbol{W}_{T,t}\|\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}\| = o(1).$$

For the second term on the right hand side of (B.4), we have

$$E\left\|\sum_{t=1}^{T} \left(\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}\right) \widetilde{\mathbb{B}}_t^0(1)\left(\boldsymbol{Z}_{2,t} - E\left(\boldsymbol{Z}_{2,t}|\mathcal{F}_{t-1}\right)\right)\right\| \leq M\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}\|\, E\|\boldsymbol{Z}_{2,t}\| \leq M\nu.$$

By choosing ν sufficiently small, and then it follows that $I_{T,61} = o_P(1)$. 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 $o_P(1)$. For $I_{T,64}$, we have

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

Similar to the proof of $\boldsymbol{I}_{T,4}$, by Lemma B.3, we can prove that $I_{T,64}$ is $o_P(1)$. Then we can conclude that $I_{T,6} = o_P(1)$.

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

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

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

17

$$
E\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t})\sum_{r=1}^{\infty}\mathbb{B}_t^r(1)\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^{\top}\right)\right\|^2
$$

$$
= E\,\mathrm{tr}\left\{\sum_{t=1}^{T}\sum_{s=1}^{T}(\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t})\sum_{r,k=1}^{\infty}\mathbb{B}_t^r(1)\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^{\top}\right)\mathrm{vec}^{\top}\left(\boldsymbol{\epsilon}_s\boldsymbol{\epsilon}_{s-k}^{\top}\right)\mathbb{B}_s^{k,\top}(1)(\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,s}^{\top})\right\}
$$

$$
\leq M\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}\|^2\sum_{r=1}^{\infty}\|\mathbb{B}_t^r(1)\|^2 \leq M\left(\max_{t\geq 1}\sum_{r=1}^{\infty}\|\mathbb{B}_t^r(1)\|\right)^2\max_{t\geq 1}\|\boldsymbol{W}_{T,t}\|\sum_{t=1}^{T}\|\boldsymbol{W}_{T,t}\| = o(1).
$$

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 $o_P(1)$.

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 complete. $\qquad\square$

Proof of Lemma B.5.

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. 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 $\gamma_T = \sqrt{d_T \log T}$.

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

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

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

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

$$
:= J_{T,11} + J_{T,12} + J_{T,13}.
$$

18

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

$$E|J_{T,12}| \le O(1)d_T \sum_{t=1}^{T} E\left\|\epsilon_t''\right\| = O(1)d_T \sum_{t=1}^{T} E\|\epsilon_t I(\|\epsilon_t\| \ge T^{\frac{1}{\delta}})\|$$

$$\le O(1)d_T \sum_{t=1}^{T} \left\{E\|\epsilon_t\|^\delta\right\}^{\frac{1}{\delta}} \left\{E\|I(\|\epsilon_t\| \ge T^{\frac{1}{\delta}})\|\right\}^{\frac{\delta-1}{\delta}}$$

$$= O(1)d_T \sum_{t=1}^{T} \left\{E\|\epsilon_t\|^\delta\right\}^{\frac{1}{\delta}} \left\{\Pr(\|\epsilon_t\| \ge T^{\frac{1}{\delta}})\right\}^{\frac{\delta-1}{\delta}}$$

$$\le O(1)d_T \sum_{t=1}^{T} \left\{E\|\epsilon_t\|^\delta\right\}^{\frac{1}{\delta}} \left\{\frac{E\|\epsilon_t\|^\delta}{T}\right\}^{\frac{\delta-1}{\delta}}$$

$$= O(T^{\frac{1}{\delta}}d_T) = o\left(\sqrt{d_T \log T}\right),$$

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 $\boldsymbol{Y}_t = \boldsymbol{W}_{T,t}(s_l)\mathbb{B}_t(1)\left(\epsilon_t' - E(\epsilon_t'|\mathcal{F}_{t-1})\right)$ for $1 \le t \le T$ and $A_T = 2T^{\frac{1}{\delta}}d_T \max_{t\ge1}\|\mathbb{B}_t(1)\|$. Simple algebra shows that $\|\boldsymbol{Y}_t\| \le A_T$ uniformly in t and s_l. By Assumption 2 and the first condition in the body of this lemma,

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

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

$$\Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T\right)$$

$$= \Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T, \; \max_{1\le l\le N_T} \left\|\sum_{t=1}^{T} E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| \le Md_T\right)$$

$$+ \Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T, \; \max_{1\le l\le N_T} \left\|\sum_{t=1}^{T} E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| > Md_T\right)$$

$$\le \Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T, \; \max_{1\le l\le N_T} \left\|\sum_{t=1}^{T} E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| \le Md_T\right)$$

$$+ \Pr\left(\max_{1\le l\le N_T} \left\|\sum_{t=1}^{T} E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| > Md_T\right)$$

$$\le N_T \exp\left(-\frac{\beta M\gamma_T^2}{2(Md_T + \gamma_T 2A_T)}\right) + 0$$

$$\le N_T \exp\left(-\frac{\beta}{2}\log T\right) = O(\delta_T^{-1})T^{-\frac{\beta}{2}} \to 0.$$

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

Proof of Lemma B.6.

(1). Similar to the proof of Lemma B.5, 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 $\gamma_T = \sqrt{d_T \log T}$.

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

$$\leq \max_{1 \leq l \leq N_T} \left\| \sum_{t=1}^{T} (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(s_l)) \mathbb{B}_t^0(1) \left(\mathrm{vec}\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top \right) - \mathrm{vec}(\boldsymbol{I}_d) \right) \right\|$$

$$+ \max_{1 \leq l \leq N_T} \sup_{\tau \in S_l} \left\| \sum_{t=1}^{T} (\boldsymbol{I}_d \otimes (\boldsymbol{W}_{T,t}(\tau) - \boldsymbol{W}_{T,t}(s_l))) \mathbb{B}_t^0(1) \left(\mathrm{vec}\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top \right) - \mathrm{vec}(\boldsymbol{I}_d) \right) \right\|$$

$$:= J_{T,1} + J_{T,2}.$$

Start from $J_{T,2}$. Similar to the proof of Lemma B.5, since

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

by Assumption 1, we have

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

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

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

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

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

$$:= J_{T,11} + J_{T,12} + J_{T,13}.$$

As in the proof of Lemma B.5, 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 $E(\boldsymbol{Y}_t | \mathcal{F}_{t-1}) = 0$ and $\|\boldsymbol{Y}_t\| \leq 2T^{2/\delta} d_T \max_t \|\mathbb{B}_t^0(1)\|$. Since $\max_{t \geq 1} E\left(\|\boldsymbol{\epsilon}_t\|^4 | \mathcal{F}_{t-1} \right) < \infty$ a.s., we can write

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

Similar to Lemma B.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

$$\Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T\right) = \Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T, \max_{1 \le l \le N_T}\left\|\sum_{t=1}^{T} E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| \le Md_T\right)$$

$$+ \Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T, \max_{1 \le l \le N_T}\left\|\sum_{t=1}^{T} E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| > Md_T\right)$$

$$\le \Pr\left(J_{T,11} > \sqrt{\beta M}\gamma_T, \max_{1 \le l \le N_T}\left\|\sum_{t=1}^{T} E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| \le Md_T\right)$$

$$+ \Pr\left(\max_{1 \le l \le N_T}\left\|\sum_{t=1}^{T} E(\boldsymbol{Y}_t\boldsymbol{Y}_t^\top|\mathcal{F}_{t-1})\right\| > Md_T\right)$$

$$\le N_T \exp\left(-\frac{\beta M\gamma_T^2}{2(Md_T + M\gamma_T T^{\frac{2}{\delta}}d_T)}\right) + 0$$

$$\le N_T \exp\left(-\frac{\beta}{2}\log T\right) = N_T T^{-\frac{\beta}{2}} = o(1).$$

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

$$\sup_{\tau \in [a,b]}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(\tau))\,\boldsymbol{\zeta}_t\boldsymbol{\epsilon}_t\right\| \le \max_{1 \le l \le N_T}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(s_l))\,\boldsymbol{\zeta}_t\boldsymbol{\epsilon}_t\right\|$$

$$+ \max_{1 \le l \le N_T}\sup_{\tau \in S_l}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d \otimes (\boldsymbol{W}_{T,t}(\tau) - \boldsymbol{W}_{T,t}(s_l)))\,\boldsymbol{\zeta}_t\boldsymbol{\epsilon}_t\right\|$$

$$:= J_{T,3} + J_{T,4}.$$

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

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

$$\le \left(E\left\|\sum_{r=1}^{\infty}\left(\sum_{s=0}^{\infty}\boldsymbol{B}_{s+r,t} \otimes \boldsymbol{B}_{s,t}\right)\text{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^\top\right)\right\|^2\right)^{1/2}$$

$$\le \left\{\text{tr}\left[\sum_{r=1}^{\infty}\left(\sum_{s=0}^{\infty}\boldsymbol{B}_{s+r,t} \otimes \boldsymbol{B}_{s,t}\right) \cdot (\boldsymbol{I}_d \otimes \boldsymbol{I}_d) \cdot \left(\sum_{s=0}^{\infty}\boldsymbol{B}_{s+r,t} \otimes \boldsymbol{B}_{s,t}\right)^\top\right]\right\}^{1/2}$$

$$\le M\left(\sum_{r=1}^{\infty}\left\|\sum_{s=0}^{\infty}\boldsymbol{B}_{s+r,t} \otimes \boldsymbol{B}_{s,t}\right\|^2\right)^{1/2}$$

$$\le M\left(\sum_{r=1}^{\infty}\left(\sum_{s=0}^{\infty}\|\boldsymbol{B}_{s+r,t}\|^2\right) \cdot \left(\sum_{s=0}^{\infty}\|\boldsymbol{B}_{s,t}\|^2\right)\right)^{1/2}$$

$$= M\left(\left(\sum_{r=1}^{\infty}r\,\|\boldsymbol{B}_{r,t}\|^2\right) \cdot \left(\sum_{s=0}^{\infty}\|\boldsymbol{B}_{s,t}\|^2\right)\right)^{1/2}$$

$$\le M\left(\left(\sum_{r=1}^{\infty}r\,\|\boldsymbol{B}_{r,t}\|\right)^2 \cdot \left(\sum_{s=0}^{\infty}\|\boldsymbol{B}_{s,t}\|^2\right)\right)^{1/2} < \infty.$$

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

$$\max_{1 \leq l \leq N_T} \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\| E\left(\|\boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t\|^2 \,|\mathcal{F}_{t-1}\right) = O_P(1). \tag{B.5}$$

Note that

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

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

$$+ \max_{1 \leq l \leq N_T} \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\| E\|\boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t\|^2$$

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.5), 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

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

$$= \max_{1 \leq l \leq N_T} \left| \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\| \operatorname{tr}\left(\sum_{r,r^*=1}^{\infty} \mathbb{B}_t^r(1)(\boldsymbol{\epsilon}_{t-r}\boldsymbol{\epsilon}_{t-r^*}^\top \otimes \boldsymbol{I}_d)\mathbb{B}_t^{r^*,\top}(1) - \sum_{r=1}^{\infty} \mathbb{B}_t^r(1)\mathbb{B}_t^{r,\top}(1) \right) \right|$$

$$\leq d^2 \cdot \max_{1 \leq l \leq N_T} \left\| \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\| \sum_{r=1}^{\infty} (\mathbb{B}_t^r(1) \otimes \mathbb{B}_t^r(1)) \left(\operatorname{vec}\left(\boldsymbol{\epsilon}_{t-r}\boldsymbol{\epsilon}_{t-r}^\top \otimes \boldsymbol{I}_d\right) - \operatorname{vec}(\boldsymbol{I}_{d^2}) \right) \right\|$$

$$+ 2d^2 \cdot \max_{1 \leq l \leq N_T} \left\| \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(s_l)\| \sum_{r=1}^{\infty} \sum_{j=1}^{\infty} \left(\mathbb{B}_t^{r+j}(1) \otimes \mathbb{B}_t^r(1) \right) \operatorname{vec}\left(\boldsymbol{\epsilon}_{t-r}\boldsymbol{\epsilon}_{t-r-j}^\top \otimes \boldsymbol{I}_d\right) \right\|$$

$$:= J_{T,5} + J_{T,6}.$$

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 B.3, we have

$$\mathbb{F}_{r,t}(L) = \mathbb{F}_{r,t}(1) - (1-L)\widetilde{\mathbb{F}}_{r,t}(L) \tag{B.6}$$

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

$$\widetilde{X}_{at} = \sum_{j=1}^{\infty} \left(\mathbb{B}_t^j(1) \otimes \mathbb{B}_t^j(1) \right) \operatorname{vec}\left(\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-j}^\top \otimes \boldsymbol{I}_d\right),$$

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

Applying (B.6) to $\widetilde{X}_{at}$ and $\widetilde{X}_{bt}$ yields that

$$\widetilde{X}_{at} = \mathbb{F}_{0,t}(1)\text{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top \otimes \boldsymbol{I}_d\right) - (1-L)\widetilde{\mathbb{F}}_{0,t}(L)\text{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top \otimes \boldsymbol{I}_d\right),$$

$$\widetilde{X}_{bt} = \sum_{r=1}^\infty \mathbb{F}_{r,t}(1)\text{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^\top \otimes \boldsymbol{I}_d\right) - (1-L)\sum_{r=1}^\infty \widetilde{\mathbb{F}}_{r,t}(L)\text{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^\top \otimes \boldsymbol{I}_d\right).$$

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

$$\max_{1 \le l \le N_T}\left\|\sum_{t=1}^T \|\boldsymbol{W}_{T,t}(s_l)\|\sum_{r=1}^\infty (\mathbb{B}_t^r(1) \otimes \mathbb{B}_t^r(1))\left(\text{vec}\left(\boldsymbol{\epsilon}_{t-r}\boldsymbol{\epsilon}_{t-r}^\top \otimes \boldsymbol{I}_d\right) - \text{vec}(\boldsymbol{I}_{d^2})\right)\right\|$$

$$\le \max_{1 \le l \le N_T}\left\|\sum_{t=1}^T \|\boldsymbol{W}_{T,t}(s_l)\|\,\mathbb{F}_{0,t}(1)\left(\text{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top \otimes \boldsymbol{I}_d\right) - \text{vec}\left(\boldsymbol{I}_{d^2}\right)\right)\right\|$$

$$+ \max_{1 \le l \le N_T}\left\|\|\boldsymbol{W}_{T,1}(s_l)\|\widetilde{\mathbb{F}}_{0,1}(L)\text{vec}\left(\boldsymbol{\epsilon}_0\boldsymbol{\epsilon}_0^\top \otimes \boldsymbol{I}_d\right)\right\| + \sup_{0 \le \tau \le 1}\left\|\|\boldsymbol{W}_{T,T}(s_l)\|\widetilde{\mathbb{F}}_{0,T}(L)\text{vec}\left(\boldsymbol{\epsilon}_T\boldsymbol{\epsilon}_T^\top \otimes \boldsymbol{I}_d\right)\right\|$$

$$+ \max_{1 \le l \le N_T}\left\|\sum_{t=1}^{T-1}\left(\|\boldsymbol{W}_{T,t+1}(s_l)\|\widetilde{\mathbb{F}}_{0,t+1}(L) - \|\boldsymbol{W}_{T,t}(s_l)\|\widetilde{\mathbb{F}}_{0,t}(L)\right)\text{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top \otimes \boldsymbol{I}_d\right)\right\|$$

$$:= J_{T,51} + J_{T,52} + J_{T,53} + J_{T,54}.$$

Similar to the proof of Lemma B.6.1, we can show that $J_{T,51} = O_P\left(\sqrt{d_T \log T}\right)$, since

$$\max_{t \ge 1}\|\mathbb{F}_{0,t}(1)\| \le \max_{t \ge 1}\sum_{j=1}^\infty\left\|\sum_{k=0}^\infty \boldsymbol{B}_{k+j,t} \otimes \boldsymbol{B}_{k,t}\right\|^2 \le \max_{t \ge 1}\sum_{j=1}^\infty\left(\sum_{k=0}^\infty \|\boldsymbol{B}_{k+j,t}\|^2\right)\left(\sum_{k=0}^\infty \|\boldsymbol{B}_{k,t}\|^2\right)$$

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

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

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

$$\le \max_{t \ge 1}\left(\sum_{j=0}^\infty \|\boldsymbol{B}_{j,t}\|^2\right)\left(\sum_{r=1}^\infty\sum_{k=r+1}^\infty (k-r)\,\|\boldsymbol{B}_{k,t}\|^2\right)$$

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

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

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

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

23

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

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

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

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

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

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

$$\sum_{t=1}^{T-1} \sum_{r=1}^{\infty} \left\| \widetilde{\mathbb{F}}_{r,t+1}(1) - \widetilde{\mathbb{F}}_{r,t}(1) \right\|$$

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

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

$$\leq \left(\sum_{t=1}^{T-1} \sum_{r=1}^{\infty} \left\| \mathbb{B}_{t+1}^r(1) - \mathbb{B}_t^r(1) \right\| \right) \left(\max_{t \geq 1} \sum_{j=1}^{\infty} j \left\| \mathbb{B}_t^j \right\| \right)$$

$$+ \left(\sum_{t=1}^{T-1} \sum_{j=1}^{\infty} j \left\| \mathbb{B}_{t+1}^j(1) - \mathbb{B}_t^j(1) \right\| \right) \left(\max_{t \geq 1} \sum_{r=1}^{\infty} \|\mathbb{B}_t^r\| \right)$$

$$= O(1).$$

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

$$J_{T,61} = \max_{1 \leq l \leq N_T} \left\| \sum_{t=1}^T \|\boldsymbol{W}_{T,t}(s_l)\| \left(\boldsymbol{u}_t' + \boldsymbol{u}_t'' - E(\boldsymbol{u}_t' + \boldsymbol{u}_t''|\mathcal{F}_{t-1}) \right) \right\|$$

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

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

$$:= J_{T,611} + J_{T,612} + J_{T,613}.$$

Using an argument as in the proof for $J_{T,12}$ of Lemma B.5, 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 $E(\boldsymbol{Y}_t|\mathcal{F}_{t-1}) = 0$ 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 $\max_{1 \leq l \leq N_T} \left\| \sum_{t=1}^{T} E(\boldsymbol{Y}_t \boldsymbol{Y}_t^{\top} | \mathcal{F}_{t-1}) \right\| = O_P(d_T T^{\frac{2}{\delta}})$. By Lemma B.2, and choosing $\beta = 4$, we have

$$\Pr \left(J_{T,611} > \sqrt{\beta M} \sqrt{d_T T^{\frac{2}{\delta}} \log T} \right)$$

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

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

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

$$+ \Pr \left(\max_{1 \leq l \leq N_T} \left\| \sum_{t=1}^{T} E(\boldsymbol{Y}_t \boldsymbol{Y}_t^{\top} | \mathcal{F}_{t-1}) \right\| > M d_T T^{\frac{2}{\delta}} \right)$$

$$\leq N_T \exp \left(-\frac{\beta M d_T T^{\frac{2}{\delta}} \log T}{2(M d_T T^{\frac{2}{\delta}} + M \sqrt{d_T T^{\frac{2}{\delta}} \log T} T^{\frac{2}{\delta}} d_T)} \right) + o(1)$$

$$\leq N_T \exp \left(-\frac{\beta}{2} \log T \right) = N_T T^{-\frac{\beta}{2}} = o(1)$$

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 $\boldsymbol{u}_t = \boldsymbol{\zeta}_t \boldsymbol{\epsilon}_t$, $\boldsymbol{u}_t' = \boldsymbol{u}_t I \left(\|\boldsymbol{u}_t\| \leq T^{\frac{2}{\delta}} \right)$ and $\boldsymbol{u}_t'' = \boldsymbol{u}_t - \boldsymbol{u}_t'$. Then we have

$$J_{T,3} = \max_{1 \leq l \leq N_T} \left\| \sum_{t=1}^{T} (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(s_l)) \left(\boldsymbol{u}_t' + \boldsymbol{u}_t'' - E(\boldsymbol{u}_t' + \boldsymbol{u}_t'' | \mathcal{F}_{t-1}) \right) \right\|$$

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

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

$$= J_{T,31} + J_{T,32} + J_{T,33}.$$

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 $1 \leq l \leq N_T$, let $\boldsymbol{Y}_t = (\boldsymbol{I}_d \otimes \boldsymbol{W}_{T,t}(s_l))(\boldsymbol{u}_t' - E(\boldsymbol{u}_t' | \mathcal{F}_{t-1}))$, then we have $E(\boldsymbol{Y}_t | \mathcal{F}_{t-1}) = 0$ and

$\|\boldsymbol{Y}_t\| \leq 2T^{2/\delta}d_T$. Also,

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

which yields that

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

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

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

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

We now have completed the proof of the second result. $\qquad\square$

Proof of Lemma B.7.

(1). For any fixed $\tau \in (0,1)$, let $\boldsymbol{W}_{T,t} = \boldsymbol{I_d}\frac{1}{T}\left(\frac{\tau_t - \tau}{h}\right)^k K_h(\tau_t - \tau)$. It is straightforward to verify the conditions imposed on $\boldsymbol{W}_{T,t}$, then the first result follows from Lemma B.4 immediately.

(2)–(3). Let $\boldsymbol{W}_{T,t}(\tau) = \boldsymbol{I_d}\frac{1}{T}\left(\frac{\tau_t - \tau}{h}\right)^k K_h(\tau_t - \tau)$, then the second and third results follow from Theorem 2.1.

$\qquad\square$

Proof of Lemma B.8.

(1). This proof is similar to Lemma B.6 with $\boldsymbol{Z}_{t-1}$ replacing $\boldsymbol{\zeta}_t$, so omitted here.

(2). Note that

$$\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} \boldsymbol{\eta}_t\left(\boldsymbol{\eta}_t - \widehat{\boldsymbol{\eta}}_t\right)^\top K\left(\frac{\tau_t - \tau}{h}\right)$$

$$= \frac{1}{\sqrt{Th}} \sum_{t=1}^{T} \boldsymbol{\eta}_t \boldsymbol{z}_{t-1}^{\top} \left(\widehat{\boldsymbol{A}}(\tau_t) - \boldsymbol{A}(\tau_t) \right)^{\top} K \left(\frac{\tau_t - \tau}{h} \right)$$

$$= \frac{1}{\sqrt{Th}} \sum_{t=1}^{T} \boldsymbol{\eta}_t \boldsymbol{z}_{t-1}^{\top} h \boldsymbol{S}_{T,0}^{-1}(\tau_t) \boldsymbol{S}_{T,1}(\tau_t) \boldsymbol{A}^{(1),\top}(\tau_t) K \left(\frac{\tau_t - \tau}{h} \right)$$

$$+ \frac{1}{\sqrt{Th}} \sum_{t=1}^{T} \boldsymbol{\eta}_t \boldsymbol{z}_{t-1}^{\top} \frac{1}{2} h^2 \boldsymbol{S}_{T,0}^{-1}(\tau_t) \boldsymbol{S}_{T,2}(\tau_t) \boldsymbol{A}^{(2),\top}(\tau_t) K \left(\frac{\tau_t - \tau}{h} \right)$$

$$+ \frac{1}{\sqrt{Th}} \sum_{t=1}^{T} \boldsymbol{\eta}_t \boldsymbol{z}_{t-1}^{\top} \boldsymbol{S}_{T,0}^{-1}(\tau_t) \left(\frac{1}{Th} \sum_{s=1}^{T} \boldsymbol{z}_{s-1} \boldsymbol{z}_{s-1}^{\top} \boldsymbol{M}^{\top}(\tau_s) K \left(\frac{\tau_s - \tau_t}{h} \right) \right) K \left(\frac{\tau_t - \tau}{h} \right)$$

$$+ \frac{1}{\sqrt{Th}} \sum_{t=1}^{T} \boldsymbol{\eta}_t \boldsymbol{z}_{t-1}^{\top} \boldsymbol{S}_{T,0}^{-1}(\tau_t) \left(\frac{1}{Th} \sum_{s=1}^{T} \boldsymbol{z}_{s-1} \boldsymbol{\eta}_s^{\top} K \left(\frac{\tau_s - \tau_t}{h} \right) \right) K \left(\frac{\tau_t - \tau}{h} \right)$$

$$:= J_{T,1} + J_{T,2} + J_{T,3} + J_{T,4}.$$

For $J_{T,1}$ to $J_{T,3}$, using Lemma B.7, we can replace the sample covariance matrix with its converged value with rate $O_P \left(\sqrt{\frac{\log T}{Th}} \right)$ and hence it's easy to show that $J_{T,1}$ to $J_{T,3}$ are $o_P(1)$.

For $J_{T,4}$, for notational simplicity, we ignore $\boldsymbol{S}_{T,0}^{-1}(\tau_t)$ and hence,

$$J_{T,4} = \frac{1}{(Th)^{3/2}} \sum_{t=1}^{T} \boldsymbol{\eta}_t \boldsymbol{z}_{t-1}^{\top} \boldsymbol{z}_{t-1} \boldsymbol{\eta}_t^{\top} K(0) K \left(\frac{\tau_t - \tau}{h} \right)$$

$$+ \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \boldsymbol{z}_{t-1}^{\top} \boldsymbol{z}_{t+i-1} \boldsymbol{\eta}_{t+i}^{\top} K \left(\frac{i}{Th} \right) K \left(\frac{\tau_t - \tau}{h} \right)$$

$$+ \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_{t+i} \boldsymbol{z}_{t+i-1}^{\top} \boldsymbol{z}_{t-1} \boldsymbol{\eta}_t^{\top} K \left(\frac{i}{Th} \right) K \left(\frac{\tau_t - \tau}{h} \right)$$

$$:= J_{T,41} + J_{T,42} + J_{T,43}.$$

It's easy to see $J_{T,41} = O_P \left((Th)^{-1/2} \right)$. For $J_{T,42}$,

$$J_{T,42} = \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t E \left(\boldsymbol{z}_{t-1}^{\top} \boldsymbol{z}_{t+i-1} \right) \boldsymbol{\eta}_{t+i}^{\top} K \left(\frac{i}{Th} \right) K \left(\frac{\tau_t - \tau}{h} \right)$$

$$+ \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\boldsymbol{z}_{t-1}^{\top} \boldsymbol{z}_{t+i-1} - E \left(\boldsymbol{z}_{t-1}^{\top} \boldsymbol{z}_{t+i-1} \right) \right) \boldsymbol{\eta}_{t+i}^{\top} K \left(\frac{i}{Th} \right) K \left(\frac{\tau_t - \tau}{h} \right)$$

$$:= J_{T,421} + J_{T,422}.$$

For $J_{T,421}$,

$$E \left\| J_{T,421} \right\|^2 \leq \frac{1}{(Th)^3} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \left\{ E \left(\boldsymbol{z}_{t-1}^{\top} \boldsymbol{z}_{t+i-1} \right) E \left(\boldsymbol{\eta}_t^{\top} \boldsymbol{\eta}_t \right) E \left(\boldsymbol{\eta}_{t+i}^{\top} \boldsymbol{\eta}_{t+i} \right) \right\}^2 K^2 \left(\frac{i}{Th} \right) K^2 \left(\frac{\tau_t - \tau}{h} \right)$$

$$= O \left(\frac{1}{Th} \right),$$

which then yields that $J_{T,421} = O_P\left((Th)^{-1/2}\right)$. For $J_{T,422}$,

$$J_{T,422} = \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \sum_{m=1}^{p} \left(\boldsymbol{x}_{t-m}^\top \boldsymbol{x}_{t+i-m} - E\left(\boldsymbol{x}_{t-m}^\top \boldsymbol{x}_{t+i-m}\right)\right) \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right).$$

For notational simplicity, assume $p = 1$ and thus

$$\begin{aligned}
J_{T,422} &= \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\boldsymbol{\mu}_{t-1}^\top \sum_{j=0}^{\infty} \boldsymbol{\Psi}_{j,t+i-1} \boldsymbol{\eta}_{t+i-1-j} \right) \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right) \\
&+ \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\sum_{j=0}^{\infty} \boldsymbol{\eta}_{t-1-j}^\top \boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\mu}_{t+i-1} \right) \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right) \\
&+ \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\sum_{j=0}^{\infty} \left(\boldsymbol{\eta}_{t-1-j}^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top - E\left(\boldsymbol{\eta}_{t-1-j}^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top\right)\right) \right. \\
&\quad \left. \cdot \operatorname{vec}\left(\boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\Psi}_{j+i,t+i-1}\right)\right) \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right) \\
&+ \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\sum_{j=0}^{\infty} \sum_{m=0, \neq j+i}^{\infty} \left(\boldsymbol{\eta}_{t+i-1-m}^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top\right) \operatorname{vec}\left(\boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\Psi}_{m,t+i-1}\right)\right) \\
&\quad \cdot \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right) \\
&:= J_{T,4221} + J_{T,4222} + J_{T,4223} + J_{T,4224}.
\end{aligned}$$

For $J_{T,4221}$,

$$\begin{aligned}
& E \left\| J_{T,4221} \right\|^2 \\
&\leq \frac{1}{(Th)^3} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} E \left\| \boldsymbol{\eta}_t \left(\boldsymbol{\mu}_{t-1}^\top \sum_{j=0}^{\infty} \boldsymbol{\Psi}_{j,t+i-1} \boldsymbol{\eta}_{t+i-1-j} \right) \right\|^2 E \left\| \boldsymbol{\eta}_{t+i} \right\|^2 K^2\left(\frac{i}{Th}\right) K^2\left(\frac{\tau_t - \tau}{h}\right) \\
&= O\left(\frac{1}{Th}\right).
\end{aligned}$$

Similarly, $J_{T,4222}$ and $J_{T,4223}$ are $O_P\left((Th)^{-1/2}\right)$. For $J_{T,4224}$,

$$\begin{aligned}
J_{T,4224} &= \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\sum_{j=0}^{\infty} \left(\boldsymbol{\eta}_t^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top \right) \operatorname{vec}\left(\boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\Psi}_{i-1,t+i-1}\right)\right) \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right) \\
&+ \frac{1}{(Th)^{3/2}} \sum_{i=1}^{T-1} \sum_{t=1}^{T-i} \boldsymbol{\eta}_t \left(\sum_{j=0}^{\infty} \sum_{m=0, \neq j+i, \neq i-1}^{\infty} \left(\boldsymbol{\eta}_{t+i-1-m}^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top\right) \right. \\
&\quad \left. \cdot \operatorname{vec}\left(\boldsymbol{\Psi}_{j,t-1}^\top \boldsymbol{\Psi}_{m,t+i-1}\right)\right) \boldsymbol{\eta}_{t+i}^\top K\left(\frac{i}{Th}\right) K\left(\frac{\tau_t - \tau}{h}\right) \\
&:= J_{T,42241} + J_{T,42242}.
\end{aligned}$$

Similar to the proof of $J_{T,4221}$, we can show that $J_{T,42242} = O_P\left((Th)^{-1/2}\right)$.

Let $\boldsymbol{w}_{t,i} = \sum_{j=0}^{\infty}\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top \otimes \boldsymbol{\eta}_{t-1-j}^\top\right)\mathrm{vec}\left(\boldsymbol{\Psi}_{j,t-1}^\top\boldsymbol{\Psi}_{i-1,t+i-1}^\top\right)$. For $J_{T,42241}$,

$$E\left\|J_{T,42241}\right\|^2 = \frac{1}{(Th)^3}\sum_{i_1=1}^{T-1}\sum_{i_2=1}^{T-1}\sum_{t_1=1}^{T-i_1}E\left\|\boldsymbol{\eta}_{t_1+i_1}^\top\boldsymbol{\eta}_{t_1+i_1}\right\|E\left\|\boldsymbol{w}_{t_1+i_1-i_2,i_2}^\top\boldsymbol{w}_{t_1,i_1}\right\|$$

$$\cdot K\left(\frac{i_1}{Th}\right)K\left(\frac{i_2}{Th}\right)K\left(\frac{\tau_{t_1}-\tau}{h}\right)K\left(\frac{\tau_{t_1+i_1-i_2}-\tau}{h}\right)$$

$$\leq \frac{M}{(Th)^3}\sum_{t_1=1}^{T}\left(\max_t\sum_{i=1}^{T-1}\|\boldsymbol{\Psi}_{i,t}\|\right)^2\left(\max_t\sum_{j=0}^{\infty}\|\boldsymbol{\Psi}_{j,t}\|\right)^2 K\left(\frac{\tau_{t_1}-\tau}{h}\right)$$

$$= O\left((Th)^{-2}\right).$$

Hence, $J_{T,42} = O_P\left((Th)^{-1/2}\right)$. Similar to $J_{T,42}$, $J_{T,43} = O_P\left((Th)^{-1/2}\right)$. The proof is now complete. $\qquad\square$

Define $\boldsymbol{\Lambda}_{\mathsf{p}}(\tau) = [\boldsymbol{a}(\tau), \boldsymbol{A}_{\mathsf{p},1}(\tau), ..., \boldsymbol{A}_{\mathsf{p},\mathsf{p}}(\tau)]$, where $\boldsymbol{A}_{\mathsf{p},j}(\tau) = \boldsymbol{A}_j(\tau)$ for $1 \leq j \leq p$ and $\boldsymbol{A}_{\mathsf{p},j}(\tau) = 0$ for $j > p$. Let $\boldsymbol{z}_{\mathsf{p},t-1} = [1, \boldsymbol{x}_{t-1}^\top, ..., \boldsymbol{x}_{t-\mathsf{p}}^\top]^\top$, $\boldsymbol{M}_{\mathsf{p}}(\tau_t) = \boldsymbol{\Lambda}_{\mathsf{p}}(\tau_t) - \boldsymbol{\Lambda}_{\mathsf{p}}(\tau) - \boldsymbol{\Lambda}_{\mathsf{p}}^{(1)}(\tau)(\tau_t-\tau) - \frac{1}{2}h^2\boldsymbol{\Lambda}_{\mathsf{p}}^{(2)}(\tau)(\tau_t-\tau)^2$, $\boldsymbol{\Lambda}_{\overline{\mathsf{p}}}(\tau) = [\boldsymbol{A}_{\mathsf{p},p+1}(\tau), ..., \boldsymbol{A}_{\mathsf{p},\mathsf{P}}(\tau)]$ and $\boldsymbol{z}_{\overline{\mathsf{p}},t-1} = [\boldsymbol{x}_{t-\mathsf{p}-1}^\top, ..., \boldsymbol{x}_{t-\mathsf{P}}^\top]^\top$.

Proof of Lemma B.9.

(1). Since $\mathsf{p} \geq p$, we have $\widehat{\boldsymbol{\eta}}_{\mathsf{p},t} = \boldsymbol{\eta}_t + \left(\boldsymbol{\Lambda}_{\mathsf{p}}(\tau_t) - \widehat{\boldsymbol{\Lambda}}_{\mathsf{p}}(\tau_t)\right)\boldsymbol{z}_{\mathsf{p},t-1}$ and

$$\mathrm{RSS}(\mathsf{p}) = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t + \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{\mathsf{p},t-1}^\top\left(\boldsymbol{\Lambda}_{\mathsf{p}}(\tau_t) - \widehat{\boldsymbol{\Lambda}}_{\mathsf{p}}(\tau_t)\right)^\top\left(\boldsymbol{\Lambda}_{\mathsf{p}}(\tau_t) - \widehat{\boldsymbol{\Lambda}}_{\mathsf{p}}(\tau_t)\right)\boldsymbol{z}_{\mathsf{p},t-1}$$

$$-2\frac{1}{T}\sum_{t=1}^{T}\mathrm{tr}\left(\boldsymbol{\eta}_t\left(\boldsymbol{\eta}_t - \widehat{\boldsymbol{\eta}}_{\mathsf{p},t}\right)^\top\right)$$

$$:= \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t + I_{T,1} + I_{T,2}.$$

Since $\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t$ is m.d.s., we have $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t = \frac{1}{T}\sum_{t=1}^{T}E(\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t) + T^{-1/2}$. By Theorem 3.1,

$$I_{T,1} \leq \frac{1}{T}\sum_{t=1}^{\lfloor Th\rfloor}\|\boldsymbol{z}_{\mathsf{p},t-1}\|^2 \cdot \left\|\widehat{\boldsymbol{\Lambda}}_{\mathsf{p}}(\tau_t) - \boldsymbol{\Lambda}_{\mathsf{p}}(\tau_t)\right\|^2 + \frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T-\lfloor Th\rfloor}\|\boldsymbol{z}_{\mathsf{p},t-1}\|^2 \cdot \left\|\widehat{\boldsymbol{\Lambda}}_{\mathsf{p}}(\tau_t) - \boldsymbol{\Lambda}_{\mathsf{p}}(\tau_t)\right\|^2$$

$$+\frac{1}{T}\sum_{t=T-\lfloor Th\rfloor+1}^{T}\|\boldsymbol{z}_{\mathsf{p},t-1}\|^2 \cdot \left\|\widehat{\boldsymbol{\Lambda}}_{\mathsf{p}}(\tau_t) - \boldsymbol{\Lambda}_{\mathsf{p}}(\tau_t)\right\|^2$$

$$\leq \sup_{0\leq\tau\leq h}\left\|\widehat{\boldsymbol{\Lambda}}_{\mathsf{p}}(\tau_t) - \boldsymbol{\Lambda}_{\mathsf{p}}(\tau)\right\|^2 \cdot \frac{1}{T}\sum_{t=1}^{\lfloor Th\rfloor}\|\boldsymbol{z}_{\mathsf{p},t-1}\|^2 + \sup_{h\leq\tau\leq 1-h}\left\|\widehat{\boldsymbol{\Lambda}}_{\mathsf{p}}(\tau_t) - \boldsymbol{\Lambda}_{\mathsf{p}}(\tau)\right\|^2 \cdot \frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{T-\lfloor Th\rfloor}\|\boldsymbol{z}_{\mathsf{p},t-1}\|^2$$

$$+ \sup_{1-h\leq\tau\leq 1}\left\|\widehat{\boldsymbol{\Lambda}}_{\mathsf{p}}(\tau_t) - \boldsymbol{\Lambda}_{\mathsf{p}}(\tau)\right\|^2 \cdot \frac{1}{T}\sum_{t=T-\lfloor Th\rfloor+1}^{T}\|\boldsymbol{z}_{\mathsf{p},t-1}\|^2$$

$$= O_P\left(2h(h + (\log T/(Th))^{1/2})^2 + (1-2h)c_T^2\right) = O_P\left((h^{3/2} + (\log T/(Th))^{1/2})^2\right).$$

For $I_{T,2}$,

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_t\left(\boldsymbol{\eta}_t-\widehat{\boldsymbol{\eta}}_{\mathsf{p},t}\right)^\top = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_t\boldsymbol{z}_{\mathsf{p},t-1}^\top\left(\widehat{\boldsymbol{\Lambda}}_{\mathsf{p}}(\tau_t)-\boldsymbol{\Lambda}_{\mathsf{p}}(\tau_t)\right)^\top$$

$$= h\cdot\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_t\boldsymbol{z}_{\mathsf{p},t-1}^\top\boldsymbol{S}_{T,0}^{-1}(\tau_t)\boldsymbol{S}_{T,1}(\tau_t)\boldsymbol{A}_{\mathsf{p}}^{(1),\top}(\tau_t)$$

$$+\frac{1}{2}h^2\cdot\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_t\boldsymbol{z}_{\mathsf{p},t-1}^\top\boldsymbol{S}_{T,0}^{-1}(\tau_t)\boldsymbol{S}_{T,2}(\tau_t)\boldsymbol{A}_{\mathsf{p}}^{(2),\top}(\tau_t)$$

$$+\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_t\boldsymbol{z}_{\mathsf{p},t-1}^\top\boldsymbol{S}_{T,0}^{-1}(\tau_t)\left(\frac{1}{Th}\sum_{s=1}^{T}\boldsymbol{z}_{\mathsf{p},s-1}\boldsymbol{z}_{s-1}^\top\boldsymbol{M}_{\mathsf{p}}^\top(\tau_s)K\left(\frac{\tau_s-\tau_t}{h}\right)\right)$$

$$+\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_t\boldsymbol{z}_{\mathsf{p},t-1}^\top\boldsymbol{S}_{T,0}^{-1}(\tau_t)\left(\frac{1}{Th}\sum_{s=1}^{T}\boldsymbol{z}_{s-1}\boldsymbol{\eta}_s^\top K\left(\frac{\tau_s-\tau_t}{h}\right)\right)$$

$$:= I_{T,3}+I_{T,4}+I_{T,5}+I_{T,6}.$$

By the uniform convergence results stated in Theorem 2.1, we replace the weighed sample covariance with its converged value plus the rate $O_P\left((\log T/(Th))^{1/2}\right)$. For $I_{T,3}$, by the fact that $\boldsymbol{S}_{T,1}(\tau) = O_P\left((\log T/(Th))^{1/2}\right)$ for $\tau\in[h,1-h]$,

$$\|I_{T,3}\| = O_P\left(T^{-\frac{1}{2}}h^{\frac{3}{2}}+h(\log T/(Th))^{1/2}\right).$$

Similarly,

$$\|I_{T,4}\|+\|I_{T,5}\| = O_P\left(T^{-\frac{1}{2}}h^2+h^2(\log T/(Th))^{1/2}\right).$$

For $I_{T,6}$, let $\boldsymbol{\Sigma}(\tau)=\mathrm{plim}_{T\to\infty}\boldsymbol{S}_{T,0}(\tau)$, we have

$$I_{T,6} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_t\boldsymbol{z}_{\mathsf{p},t-1}^\top\boldsymbol{\Sigma}^{-1}(\tau_t)\left(\frac{1}{Th}\sum_{s=1}^{T}\boldsymbol{z}_{s-1}\boldsymbol{\eta}_s^\top K\left(\frac{\tau_s-\tau_t}{h}\right)\right)+O_P\left((Th)^{-1/2}(\log T/(Th))^{1/2}\right).$$

Similar to the proof of $J_{T,4}$ in Lemma B.8, we can show

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_t\boldsymbol{z}_{\mathsf{p},t-1}^\top\boldsymbol{\Sigma}^{-1}(\tau_t)\left(\frac{1}{Th}\sum_{s=1}^{T}\boldsymbol{z}_{s-1}\boldsymbol{\eta}_s^\top K\left(\frac{\tau_s-\tau_t}{h}\right)\right)=O_P((Th)^{-1}).$$

Since $(Th)^{-1}+T^{-\frac{1}{2}}h^{\frac{3}{2}}=o\left(c_T\phi_T\right)$, result (1) follows.

(2). For $\mathsf{p}<p$, we have $\widehat{\boldsymbol{\Lambda}}_{\mathsf{p}}(\tau)-\boldsymbol{\Lambda}_{\mathsf{p}}(\tau)=\boldsymbol{B}_{\mathsf{p}}(\tau)+o_P(1)$ uniformly over $\tau\in[0,1]$, where $\boldsymbol{B}_{\mathsf{p}}(\tau)$ is a nonrandom bias term. Since $\widehat{\boldsymbol{\eta}}_{\mathsf{p},t}=\boldsymbol{\eta}_t+\left(\boldsymbol{\Lambda}_{\mathsf{p}}(\tau_t)-\widehat{\boldsymbol{\Lambda}}_{\mathsf{p}}(\tau_t)\right)\boldsymbol{z}_{\mathsf{p},t-1}+\boldsymbol{\Lambda}_{\overline{\mathsf{p}}}(\tau_t)\boldsymbol{z}_{\overline{\mathsf{p}},t-1}$, by Theorem 2.1, we have

$$\mathrm{RSS}(\mathsf{p}) = \frac{1}{T}\sum_{t=1}^{T}E\left(\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t\right)+\frac{1}{T}\sum_{t=1}^{T}\mathrm{tr}\left([\boldsymbol{B}_{\mathsf{p}}(\tau_t),\boldsymbol{\Lambda}_{\overline{\mathsf{p}}}(\tau_t)]E\left(\boldsymbol{z}_{\mathsf{P},t-1}\boldsymbol{z}_{\mathsf{P},t-1}^\top\right)[\boldsymbol{B}_{\mathsf{p}}(\tau_t),\boldsymbol{\Lambda}_{\overline{\mathsf{p}}}(\tau_t)]^\top\right)+o_P(1).$$

Since $[\boldsymbol{B}_{\mathsf{p}}(\tau_t),\boldsymbol{\Lambda}_{\overline{\mathsf{p}}}(\tau_t)]\neq 0$ and $E\left(\boldsymbol{z}_{\mathsf{P},t-1}\boldsymbol{z}_{\mathsf{P},t-1}^\top\right)$ is a positive definite matrix, the result follows. $\square$