Econometric Theory, **0**, 2024, 1–33.
doi:10.1017/S0266466623000397

ASYMPTOTICS FOR TIME-VARYING VECTOR MA(∞) PROCESSES

YAYI YAN
Shanghai University of Finance and Economics and Monash University

JITI GAO
Monash University

BIN PENG
Monash University

This paper introduces a new class of time-varying vector moving average processes of infinite order. These processes serve dual purposes: (1) they can be used to model time-varying dependence structures, and (2) they can be used to establish asymptotic theories for multivariate time series models. To illustrate these two points, we first establish some fundamental asymptotic properties and use them to infer the trending term of a vector moving average infinity process. We then investigate a class of time-varying VARX models. Finally, we demonstrate the empirical relevance of the theoretical results using extensive simulated and real data studies.

1. INTRODUCTION

In the literature of time series analysis, moving average processes of infinite order (MA(∞)) are possibly one of the most fundamental data generating mechanisms (Beveridge and Nelson, 1981; Phillips and Solo, 1992; Hamilton, 1994, p. 48; Lütkepohl, 2005, p. 18). Notably, MA(∞) representation does not only facilitate the development of many asymptotic results (e.g., Xu and Phillips, 2008; Brüggemann, Jentsch, and Trenkler, 2016), but it is also widely used to model time series autocorrelations (e.g., Bühlmann, 1998; Friedrich, Smeekes, and Urbain, 2020) in different scenarios. To the best of our knowledge, most (if not all) studies involving MA(∞) processes work with constant parameters. Therefore, modeling heterogeneity along the time dimension is somewhat limited. However, as pointed out by Hansen (2001), dynamic processes with time-invariant coefficients may be

The authors of this paper would like to thank the Co-Editor, Yixiao Sun, and two referees for their constructive comments. Thanks also go to George Athanasopoulos, Rainer Dahlhaus, David Frazier, Oliver Linton, Gael Martin, Peter C. B. Phillips, and Wei Biao Wu for their helpful comments on earlier versions of this paper. Yan acknowledges the financial support of the National Natural Science Foundation of China (Grant No. 72303142) and Fundamental Research Funds for the Central Universities (Grant Nos. 2022110877 & 2023110099). Both Gao and Peng acknowledge the Australian Research Council Discovery Grants Program for its financial support under Grant Numbers: DP200102769 and DP210100476. Address correspondence to Jiti Gao, Department of Econometrics and Business Statistics, Monash University, Caulfield East, VIC 3145, Australia; e-mail: Jiti.Gao@monash.edu.

unnecessarily restrictive. Thus, our first goal is to marry MA(∞) representation and a nonparametric framework to accommodate dynamics better.

We now review some relevant literature. In recent years, nonparametric framework has been extensively adopted to study deterministically unknown time-varying parameters of autoregressive models (e.g., Dahlhaus and Rao, 2006; Dahlhaus, 2012; Zhang and Wu, 2012; Richter and Dahlhaus, 2019; Sun et al., 2021). Many models are often referred to as "locally stationary processes," because their data generating processes change smoothly over time. This line of research mainly deals with univariate time series and requires locally stationary mixing (e.g., Sun et al., 2021; Casini, 2023). However, stationary linear processes (including some simple AR(1) processes) are not necessarily mixing (Doukhan, 2012, Sect. 2.3.1), unless imposing some restrictive conditions on the densities of error terms (Withers, 1981). Another equally important strand of literature assumes that the coefficients of interest evolve in a random way (e.g., Primiceri, 2005; Petrova, 2019), which relies on Bayesian algorithms to produce estimation results. Despite its great popularity, asymptotic properties of these methods are barely explored (Giraitis, Kapetanios, and Yates, 2014).

This paper takes another route, showing the versatility of vector moving average (VMA(∞)) processes with nonparametrically unknown time-varying coefficients. Specifically, we start with deriving a time-varying version of Beveridge–Nelson (BN) decomposition (Beveridge and Nelson, 1981; Phillips and Solo, 1992), which yields the time-varying counterparts of long-run and transitory elements. As a sequence, time-varying versions of the (functional) central limit theorem considered in Phillips and Solo (1992) remain true. In addition, we provide several new asymptotic properties for time-varying linear processes, including the uniform consistency, the bootstrap consistency, and the long-run covariance matrix estimation.

The newly proposed framework can be used to model time-varying dependence structure, and can also serve as a device to establish asymptotic theories for many multivariate time series models. For the purpose of illustration, we consider inferring the trending term of a VMA(∞) process and investigate a class of time-varying VARX models. We then examine the aforementioned theoretical findings using extensive simulations. In an empirical study, we study the long-run level of inflation and the natural rate of unemployment using U.S. data, and find that (1) the long-run level of inflation is more anchored now and is close to the Federal Reserve's target of 2% after the Great Moderation period began, and (2) the natural rate of unemployment is less persistent and increases rapidly during the "Second Oil Crisis" and "Global Financial Crisis."

The paper is organized as follows: Section 2 introduces a class of time-varying VMA(∞) processes, develops a time-varying counterpart of the conventional BN decomposition, and establishes a set of asymptotic properties. Sections 3 and 4 apply the results of Section 2 to infer the trend of a time-varying VMA(∞) model, and study a class of time-varying VARX models. Section 5 provides extensive simulations and a real data example. Section 6 gives a short conclusion. The

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

preliminary lemmas and the proofs of some selected proofs of the main results are given in Appendix. In Appendix B of the Supplementary Material, we first discuss some impulse response analyses, and then include the omitted proofs of the main results.

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

2. THE SETUP AND ASYMPTOTICS

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

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

where $\boldsymbol{x}_t$ is a vector of d-dimensional observable variables, $\boldsymbol{\mu}_t$ is a vector of d-dimensional unknown deterministic trending functions, all $\boldsymbol{B}_{j,t}$ are $d\times d$ unknown matrices, $\boldsymbol{\epsilon}_t$ is a vector of d-dimensional random innovations, and d is fixed. Obviously, $\mathbb{B}_t(L)=\sum_{j=0}^{\infty}\boldsymbol{B}_{j,t}L^j$, where L is the lag operator.

We require the following conditions to hold throughout.

Assumption 1. Let $\mathbb{B}_t(1)$ be of full rank for each given t, $\lim_{T\to\infty}\sum_{t=1}^{T-1}\|\boldsymbol{\mu}_{t+1}-\boldsymbol{\mu}_t\|<\infty$, $\sup_{t\geq 1}\sum_{j=1}^{\infty}j\|\boldsymbol{B}_{j,t}\|<\infty$ and $\lim_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty}j\|\boldsymbol{B}_{j,t+1}-\boldsymbol{B}_{j,t}\|<\infty$.

Assumption 2. Suppose that $\{\boldsymbol{\epsilon}_t\}_{t=-\infty}^{\infty}$ is a martingale difference sequence (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 $\sup_{t\geq 1}E\,\|\boldsymbol{\epsilon}_t\|^\delta<\infty$ for some $\delta>4$.

Assumption 1 regulates the matrices $\boldsymbol{B}_{j,t}$'s, and ensures the validity of the BN decomposition under a time-varying framework. It covers many cases, including (i) the parametric setting of Phillips and Solo (1992), and (ii) $\boldsymbol{B}_{j,t}:=\boldsymbol{B}_j(t/T)$, where $\boldsymbol{B}_j(\cdot)$ satisfies Lipschitz continuity on $[0,1]$ for all j with the Lipschitz constant of order $O(j^{-(2+\nu)})$ for some $\nu>0$. These conditions can easily be verified as they are directly related to some commonly used data generating mechanisms (see, for example, Proposition 2.1). Assumption 2 is rather standard (e.g., Dahlhaus and Polonik, 2009).

For (2.1), an application of the BN decomposition immediately yields

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

where $\mathbb{B}_t(L) = \mathbb{B}_t(1) - (1-L)\widetilde{\mathbb{B}}_t(L)$, $\widetilde{\mathbb{B}}_t(L) = \sum_{j=0}^{\infty} \widetilde{\boldsymbol{B}}_{j,t} L^j$, and $\widetilde{\boldsymbol{B}}_{j,t} = \sum_{k=j+1}^{\infty} \boldsymbol{B}_{k,t}$. Consequently, the following lemma holds.

LEMMA 2.1. *Let Assumptions 1 and 2 hold. Suppose that there exists a function* $\mathbb{B}(u)$ *on* $[0,1]$*, and is continuous except for a finite number of discontinuities such that*

$$\limsup_{T\to\infty} \max_{1\le t\le T} \|\mathbb{B}_t(1) - \mathbb{B}(t/T)\| = 0.$$

As $T \to \infty$*, for* $\forall r \in [0,1]$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \left(\boldsymbol{x}_t - \boldsymbol{\mu}_t\right) \to_D \int_0^r \mathbb{B}(u) \mathrm{d}\boldsymbol{W}(u),$$

where $\boldsymbol{W}(\cdot)$ *is a standard multivariate Brownian motion.*

By Lemma 2.1, it is easy to see that (2.1) extends similar treatments by Phillips and Solo (1992) for the conventional univariate linear process, and allows one to relax many $I(0)$ and $I(1)$ related results of the literature using a time-varying VMA(∞) framework. Below, we list two examples, of which the parametric counterparts can be found in, for example, Lütkepohl (2005, pp. 387, 419).

Example 1. Suppose that $\boldsymbol{x}_t$ is a d-dimensional time-varying VARMA(p,q) process as follows:

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

where the roots of $\boldsymbol{I}_d - \boldsymbol{A}_{1,t} - \cdots - \boldsymbol{A}_{p,t} = \boldsymbol{0}_d$ all lie outside the unit circle. Simple algebra shows that (2.3) can be expressed as $\boldsymbol{x}_t = \sum_{b=0}^{\infty} \boldsymbol{D}_{b,t}\boldsymbol{\epsilon}_{t-b}$ with $\boldsymbol{D}_{b,t} = \sum_{j=\max(0,b-q)}^{b} \boldsymbol{B}_{j,t}\boldsymbol{\Theta}_{b-j,t-j}$, in which $\boldsymbol{B}_{j,t} = \boldsymbol{J}\prod_{i=0}^{j-1}\boldsymbol{\Phi}_{t-i}\boldsymbol{J}^\top$, $\boldsymbol{J} = [\boldsymbol{I}_d, \boldsymbol{0}_{d\times d(p-1)}]$, $\boldsymbol{\Phi}_t$ is the companion matrix, and $\boldsymbol{\Theta}_{0,t} \equiv \boldsymbol{I}_d$ for all t.

Example 2. Suppose that $\boldsymbol{x}_t$ is a d-dimensional time-varying VARX process of the form:

$$\boldsymbol{x}_t = \boldsymbol{A}_{1,t}\boldsymbol{x}_{t-1} + \ldots + \boldsymbol{A}_{p,t}\boldsymbol{x}_{t-p} + \boldsymbol{\Theta}_t\boldsymbol{z}_t + \boldsymbol{\epsilon}_t \quad \text{and} \quad \boldsymbol{z}_t = \sum_{j=0}^{\infty} \boldsymbol{C}_{j,t}\boldsymbol{v}_{t-j}, \tag{2.4}$$

where $\boldsymbol{z}_t$ is an m-dimensional vector and $\boldsymbol{\Theta}_t$ is a $d \times m$ matrix. Under the invertibility condition as in Example 1, model (2.4) can be further written as

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

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

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

Formally, we summarize the main results regarding Examples 1 and 2 in the following proposition.

PROPOSITION 2.1. *Consider Examples 1 and 2, and 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;*
- $\lim_{T\to\infty}\sum_{t=1}^{T-1}\|\boldsymbol{A}_{m,t+1}-\boldsymbol{A}_{m,t}\|<\infty$ *for* $m=1,\ldots,p$, *and* $\boldsymbol{A}_{m,t}=\boldsymbol{A}_{m,1}$ *for* $t\le 0$ *and* $m=1,\ldots,p$;
- $\lim_{T\to\infty}\sum_{t=1}^{T-1}\|\boldsymbol{\Theta}_{m,t+1}-\boldsymbol{\Theta}_{m,t}\|<\infty$ *for* $m=1,\ldots,q$.

1. *For Example 1, (2.3) admits a time-varying VMA(∞) process, of which the coefficients satisfy Assumption 1.*
2. *For Example 2, suppose also that* $\lim_{T\to\infty}\sum_{t=1}^{T-1}\sum_{j=1}^{\infty}j\|\boldsymbol{C}_{j,t+1}-\boldsymbol{C}_{j,t}\|<\infty$ *and* $\sup_{t\ge 1}\sum_{j=1}^{\infty}j\|\boldsymbol{C}_{j,t}\|<\infty$. *Then, (2.4) admits a time-varying VMA(∞), of which the coefficients satisfy Assumption 1.*

Notably, the assumptions of Proposition 2.1 are rather minor. For example, we may allow for structural breaks and smooth structural changes simultaneously as follows:

$$\boldsymbol{A}_{j,t}=\sum_{i=1}^{K_0}\boldsymbol{C}_{ji}(\tau_t)\,\mathrm{I}(r_{i-1}<\tau_t\le r_i),$$

where $\mathrm{I}(\cdot)$ denotes the indicator function, each $\boldsymbol{C}_{ji}(\tau)$ is defined similarly to $\boldsymbol{A}_j(\tau)$, K_0 is an unknown finite integer representing the number of breaks, and $0=r_0<r_1<\cdots<r_{K_0}<1$ are the time stamps of the change points.

By allowing for discontinuities across some regimes, we can deal with relevant features such as the Great Moderation with the gradual decline in variance for many macroeconomic variables (Stock and Watson, 2016b) or the dramatic effects of the COVID-19 pandemic (Lenza and Primiceri, 2022). The conditions on $\boldsymbol{B}_{j,t}$ of Assumption 1 are flexible enough to accommodate such structural breaks.

2.1. Asymptotic Properties for Sample Moments

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

LEMMA 2.2. *Let Assumptions 1 and 2 hold. Suppose that* $\{\boldsymbol{W}_{T,t}\}_{t=1}^{T}$ *is a sequence of* $m\times d$ *deterministic weighting matrices with m being fixed satisfying:*

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

- $\sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}\| = O(1)$;
- $\sum_{t=1}^{T-1} \left\|\boldsymbol{W}_{T,t+1} - \boldsymbol{W}_{T,t}\right\| = O(d_T)$ *with* $d_T = \sup_{t\geq 1} \|\boldsymbol{W}_{T,t}\| \to 0$.

As $T \to \infty$, *for any fixed integer* p (≥ 0),

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

Lemma 2.2 establishes convergence rates for some weighted sample moments, and covers both the parametric rate $d_T = \frac{1}{T}$ (e.g., $\boldsymbol{W}_{T,t} = \frac{1}{T}$) and nonparametric rate $d_T = \frac{1}{Th}$ (e.g., $\boldsymbol{W}_{T,t} = \frac{1}{T}K_h(\tau_t - \tau)$ for $\forall \tau \in [0,1]$).

We then strengthen the results of Lemma 2.2 with rates of uniform convergence.

LEMMA 2.3. *Let Assumptions 1 and 2 hold. Suppose that* $\left\{\boldsymbol{W}_{T,t}(\cdot)\right\}_{t=1}^{T}$ *is a sequence of* $m \times d$ *matrices of deterministic weighting functions with* m *being fixed satisfying:*

- *each functional component of* $\boldsymbol{W}_{T,t}(\cdot)$ *is Lipschitz continuous and defined on a compact set* $[a,b]$;
- $\sup_{\tau\in[a,b]} \sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}(\tau)\| = O(1)$;
- $\sup_{\tau\in[a,b]} \sum_{t=1}^{T-1} \|\boldsymbol{W}_{T,t+1}(\tau) - \boldsymbol{W}_{T,t}(\tau)\| = O(d_T)$ *with* $d_T = \sup_{\tau\in[a,b],\, t\geq 1} \|\boldsymbol{W}_{T,t}(\tau)\| \to 0$.

As $T \to \infty$, *for any fixed integer* p (≥ 0),

1. $\sup_{\tau\in[a,b]} \| \sum_{t=1}^{T} \boldsymbol{W}_{T,t}(\tau)\, (\boldsymbol{x}_t - E(\boldsymbol{x}_t)) \| = O_P(\sqrt{d_T \log T})$ *if* $T^{\frac{2}{\delta}} d_T \log T \to 0$;
2. $\sup_{\tau\in[a,b]} \| \sum_{t=1}^{T} \boldsymbol{W}_{T,t}(\tau) \left(\boldsymbol{x}_t\boldsymbol{x}_{t+p}^{\top} - E(\boldsymbol{x}_t\boldsymbol{x}_{t+p}^{\top})\right) \| = O_P(\sqrt{d_T \log T})$ *if* $T^{\frac{4}{\delta}} d_T \log T \to 0$ *and* $\sup_{t\geq 1} E(\|\boldsymbol{\epsilon}_t\|^4 | \mathcal{F}_{t-1}) < \infty$ *a.s.*

Lemma 2.3 corresponds to some existing uniform convergence results for nonparametric estimation of time series models without strict stationarity (e.g., Hansen, 2008; Gao et al., 2015; Li, Phillips, and Gao, 2016; Phillips, Li, and Gao, 2017). In Section 4, we apply this result to study a class of time-varying VARX models.

2.2. Inferences

To obtain valid inferences in practice, we establish a central limit theorem in Lemma 2.4, and then propose two methods: (1) the dependent wild bootstrap (DWB) approach and (2) the heteroscedasticity and autocorrelation consistent (HAC) covariance matrix estimation approach, to estimate the asymptotic covariance matrix involved in Lemma 2.4.

LEMMA 2.4. *Let Assumptions 1 and 2 hold. Suppose that* $\left\{\boldsymbol{W}_{T,t}\right\}_{t=1}^{T}$ *is a sequence of* $m \times d$ *deterministic weighting matrices with* m *being fixed satisfying:*

- $\sum_{t=1}^{T} \|\boldsymbol{W}_{T,t}\| = O(1)$, *where* $\sup_{t\geq 1} \|\boldsymbol{W}_{T,t}\| = O(d_T)$ *with* $d_T \to 0$;
- $\sum_{t=1}^{T-1} \left\|\boldsymbol{W}_{T,t+1} - \boldsymbol{W}_{T,t}\right\| = O(d_T)$.

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

As $T \to \infty$, $\frac{1}{\sqrt{d_T}} \sum_{t=1}^{T} \boldsymbol{W}_{T,t}(\boldsymbol{x}_t - E(\boldsymbol{x}_t)) \to_D N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{W}})$, *where* $\boldsymbol{\Sigma}_{\boldsymbol{W}} = \lim_{T\to\infty} \frac{1}{d_T} \sum_{t=1}^{T} \boldsymbol{W}_{T,t} \mathbb{B}_t(1) \mathbb{B}_t^\top(1) \boldsymbol{W}_{T,t}^\top$ *is a positive definite matrix.*

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

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

LEMMA 2.5. *Let* $l \to \infty$ *and* $l\sqrt{d_T} \to 0$. *Additionally, let* $a(\cdot)$ *be a symmetric and bounded function with bounded support* $[-1,1]$, $a(\cdot)$ *is continuous on* $[-1,1]$, $a(0) = 1$ *and* $K_a(x) = \int_{-\infty}^{\infty} a(u) e^{-iux} \mathrm{d}u \geq 0$ *for* $x \in \mathbb{R}$. *Under the conditions of Lemma 2.4, as* $T \to \infty$,

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

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

The condition on $K_a(x)$ ensures the semi-positive definiteness of the covariance matrix of $\{\xi_t^*\}_{t=1}^T$, and is necessary to generate dependent wild bootstrap samples in practice. The restrictions on $a(\cdot)$ are satisfied by a few commonly used kernels, such as the Bartlett and Parzen kernels (i.e., $a(x) = (1-|x|)\mathrm{I}(|x| \leq 1)$ and $a(x) = (1 - 6x^2 + 6|x|^3)\mathrm{I}(|x| \leq 1/2) + 2(1-|x|)^3 \mathrm{I}(1/2 \leq |x| \leq 1)$ in Andrews, 1991, p. 821). In practice, one may generate $\xi^* \equiv (\xi_1^*, \ldots, \xi_T^*)^\top \sim N(0, \Sigma_{\xi^*})$, where $\Sigma_{\xi^*} = \{a\left(\frac{t-s}{\ell}\right)\}_{T\times T}$. The normal distribution is not really necessary, but it fulfills the aforementioned conditions and is easy to implement.

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

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{W}} = \widehat{\boldsymbol{\Xi}}_0 + \sum_{i=1}^{\ell^*} \psi\left(i/\ell^*\right) (\widehat{\boldsymbol{\Xi}}_i + \widehat{\boldsymbol{\Xi}}_i^\top), \tag{2.5}$$

where $\widehat{\boldsymbol{\Xi}}_i = \frac{1}{d_T} \sum_{t=1}^{T-i} \boldsymbol{W}_{T,t} \boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top \boldsymbol{W}_{T,t+i}^\top$ for $i \geq 0$, $\boldsymbol{e}_t = \boldsymbol{x}_t - E(\boldsymbol{x}_t)$, $\psi(\cdot)$ is a kernel function, and ℓ^* is the bandwidth diverging at a relatively slow rate, in which $E(\boldsymbol{x}_t)$ is assumed to be computable at this stage. Otherwise, it will be replaced by an estimated version as in (3.2). Under some mild conditions, we establish asymptotic properties for (2.5) in the following lemma.

LEMMA 2.6. *Suppose that $\psi(\cdot)$ is a symmetric and bounded function with bounded support on* $[-1,1]$, *$\psi(\cdot)$ is continuous on* $[-1,1]$, *and* $\psi(0) = 1$. *Additionally, let $\ell^* \to \infty$ and $\ell^*\sqrt{d_T} \to 0$. Under the conditions of Lemma 2.4,* $\widehat{\mathbf{\Sigma}}_W = \mathbf{\Sigma}_W + o_P(1)$.

The conditions on ℓ^* and $\psi(\cdot)$ are standard, and are similar to those for the DWB method, except that we do not require the Fourier transform of $\psi(\cdot)$ to be nonnegative. It is worth mentioning that Casini (2023) considers HAC estimation of time-varying long-run covariance matrix for univariate locally stationary processes based on some cumulant conditions, and the author argues that the cumulant conditions can be verified using some mixing conditions. In this article, we utilize the MA structure, and do not impose any additional mixing conditions or cumulant conditions on the error process.

Up to this point, we have established a set of asymptotic properties for the VMA(∞) process (2.1). In the next section, we will apply these results to infer the trend of (2.1).

To close this section, we comment on another relevant literature—the fixed-b framework in which b denotes the ratio between the bandwidth and the sample size. The DWB and HAC approaches are concerned with the consistency of long-run covariance estimation, and a necessary condition is that the bandwidth goes to infinity but at a slower rate than the sample size. However, these first-order asymptotic results do not reflect the influence of bandwidth on the hypothesis testing. To account for the influence of the bandwidth on the hypothesis testing, the fixed-b framework is proposed in the heteroscedasticity autocorrelation robust testing context. For example, Sun, Phillips, and Jin (2008) point out that "…*the optimal bandwidth that minimizes a weighted average of type I and type II errors is larger by an order of magnitude than the bandwidth that minimizes the asymptotic mean squared error of the corresponding long-run variance estimator*."

However, extending the fixed-b framework to the time-varying framework may not be feasible. Consider a special case of the fix-b approach when $d = 1$, $\mu_t \equiv \mu$ and $W_{T,t} = 1/T$. By Lemma 2.1,

$$T^{-1/2}\sum_{t=1}^{T}(x_t - \mu) \to_D \int_0^1 \mathbb{B}(u)\mathrm{d}W(u).$$

We take the Bartlett kernel and let $b = 1$, so the long-run covariance estimator is $2T^{-1}\sum_{t=1}^{T}(T^{-1/2}S_t)^2$ with $S_t = \sum_{j=1}^{t}(x_j - \mu)$ (see Kiefer and Vogelsang 2002 for details). Using Lemma 2.1 again, we have

$$\frac{1}{\sqrt{T}}S_{\lfloor Tr\rfloor} \to_D \int_0^r \mathbb{B}(u)\mathrm{d}W(u),$$

and thus by the continuous mapping theorem, we can derive

$$\frac{2}{T}\sum_{t=1}^{T}(T^{-1/2}S_t)^2 \to_D 2\int_0^1 \left(\int_0^r \mathbb{B}(u)\mathrm{d}W(u)\right)^2 \mathrm{d}r.$$

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

Hence, $T^{-1/2}\sum_{t=1}^{T}(x_t-\mu_t)/\sqrt{2T^{-1}\sum_{t=1}^{T}(T^{-1/2}S_t)^2}$ does not converge to a pivotal distribution as in the time-invariant case since the nuisance parameter $\mathbb{B}(u)$ does not get canceled out.

3. ON DETERMINISTIC TREND

To facilitate the development, it is useful to impose more structure: $\boldsymbol{\mu}_t=\boldsymbol{\mu}(\tau_t)$ and $\boldsymbol{B}_{j,t}=\boldsymbol{B}_j(\tau_t)$, where $\tau_t=t/T$. Thus, (2.1) can be rewritten as

$$\boldsymbol{x}_t=\boldsymbol{\mu}(\tau_t)+\boldsymbol{e}_t, \tag{3.1}$$

where $\boldsymbol{e}_t=\sum_{j=0}^{\infty}\boldsymbol{B}_j(\tau_t)\boldsymbol{\epsilon}_{t-j}$. Below, we focus on the estimation of $\boldsymbol{\mu}(\tau)$ in the rest of this section before we make some comments at the end of this section.

The following assumptions are necessary.

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

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

Assumption 3 imposes smoothness conditions on the functional coefficients, which are easily verifiable and can be regarded as a special case of Assumption 1. Assumption 4 is standard in the literature of nonparametric kernel estimation (Li and Racine, 2007, p. 9). In the same requirements as for $a(\cdot)$ in Lemma 2.5, the choice of $[-1,1]$ is for simplicity only.

We estimate $\boldsymbol{\mu}(\tau)$ by

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

following the literature of trend function estimation (e.g., Bühlmann, 1998; Friedrich et al., 2020).

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

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

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

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

If $\alpha = 0$, there is no bias term involved in the asymptotic distribution of Theorem 3.1, which then falls in the usual undersmoothing scenario (Li and Racine, 2007, p. 15). In general, in order to establish valid inferences, both $\boldsymbol{\mu}_b(\tau)$ and $\tilde{v}_0 \boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\tau)$ have to be accounted for.

The DWB Procedure:

1. For $\forall \tau \in (0,1)$, let $\widetilde{\boldsymbol{\mu}}(\tau)$ be defined in the same way as in (3.2) using an oversmoothing bandwidth $\widetilde{h}$, and obtain the estimated residuals: $\widetilde{\boldsymbol{e}}_t = \boldsymbol{x}_t - \widetilde{\boldsymbol{\mu}}(\tau_t)$ for $t \geq 1$.
2. Generate $\boldsymbol{x}_t^* = \widetilde{\boldsymbol{\mu}}(\tau_t) + \boldsymbol{e}_t^*$ with $\boldsymbol{e}_t^* = \xi_t^* \widetilde{\boldsymbol{e}}_t$, where ξ_t^*'s form an l-dependent time series satisfying $E[\xi_t^*] = 0$, $E[\xi_t^{*2}] = 1$, $E\left[|\xi_t^*|^{\nu}\right] < \infty$ for some $\nu > 2$, and $E[\xi_t^* \xi_s^*] = a\left((t-s)/l\right)$ with a kernel function $a(\cdot)$ and a tuning parameter l.
3. Use $\boldsymbol{x}_t^*$'s to construct an estimator $\widehat{\boldsymbol{\mu}}^*(\tau)$ as in (3.2).
4. Repeat Steps 2 and 3 J times. Let $\boldsymbol{q}_{\alpha}(\tau)$ be the α-quantile of the J statistics $\widehat{\boldsymbol{\mu}}^*(\tau) - \widetilde{\boldsymbol{\mu}}(\tau)$, and denote the $(1-\alpha)\cdot 100\%$ confidence interval of $\widehat{\boldsymbol{\mu}}(\tau)$ as

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

Here, $\widetilde{h}$ is an oversmoothing bandwidth, as we shall require $h/\widetilde{h} \to 0$. An asymptotic property for the DWB procedure is given in Theorem 3.2.

THEOREM 3.2. *Let $l \to \infty$, $\max\{\widetilde{h}, h/\widetilde{h}\} \to 0$ and $l \cdot \max\{1/\sqrt{Th}, \widetilde{h}^4\} \to 0$. Additionally, let $a(\cdot)$ be a symmetric and bounded function with bounded support $[-1,1]$, $a(\cdot)$ is continuous on $[-1,1]$, $K(0) = 1$, and $K_a(x) = \int_{-\infty}^{\infty} a(u) e^{-iux} \mathrm{d}u \geq 0$ for $x \in \mathbb{R}$. Under the conditions of Theorem 3.1, for $\forall \tau \in (0,1)$,*

$$\sup_{\boldsymbol{w} \in \mathbb{R}^d} \left| \Pr^*\left[\sqrt{Th}\left(\widehat{\boldsymbol{\mu}}^*(\tau) - \widetilde{\boldsymbol{\mu}}(\tau)\right) \leq \boldsymbol{w}\right] - \Pr\left[\sqrt{Th}\left(\widehat{\boldsymbol{\mu}}(\tau) - \boldsymbol{\mu}(\tau)\right) \leq \boldsymbol{w}\right] \right| = o_P(1),$$

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

Theorem 3.2 shows that the confidence interval of $\boldsymbol{\mu}(\tau)$ can be recovered by the empirical quantile of $\widehat{\boldsymbol{\mu}}^*(\tau) - \widetilde{\boldsymbol{\mu}}(\tau)$. It implies that there is no need to deal with the bias in the DWB procedure, as the bootstrap draws generate a bias term identical to that in Theorem 3.1. We stress that we only consider point-wise inference for the time-varying coefficients functions, while simultaneous inference for the entire time-varying curves relies on deep Gaussian approximation theory (cf. Zhou and Wu, 2010), and is beyond the scope of this paper.

Note that the proposed kernel method can asymptotically recover $\boldsymbol{\mu}(\tau)$ even when the true form of $\boldsymbol{\mu}(\tau)$ is linear. Note that the proposed kernel method can also be applied to estimate a semiparametric version of the form: $\boldsymbol{\mu}(\tau) = \boldsymbol{\alpha}\tau + \boldsymbol{g}(\tau)$, where $\boldsymbol{\alpha}$ is a vector of unknown parameters. In this case, under the orthogonality condition: $\int_0^1 u\boldsymbol{g}(u)du = \boldsymbol{0}$, $\boldsymbol{\alpha}$ is identifiable and can be estimated by $\widehat{\boldsymbol{\alpha}} = \left(\sum_{t=1}^T \tau_t^2\right)^{-1}\left(\sum_{t=1}^T \tau_t \boldsymbol{x}_t\right)$.

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

We then estimate $\boldsymbol{g}(\tau)$ by $\widehat{\boldsymbol{g}}(\tau) = \left(\sum_{t=1}^{T} K_h(\tau_t - \tau)\right)^{-1} \sum_{t=1}^{T} (\boldsymbol{x}_t - \widehat{\boldsymbol{\alpha}}\,\tau_t) K_h(\tau_t - \tau)$, and the corresponding asymptotic properties of $\widehat{\boldsymbol{g}}(\tau)$ follow similarly from those of $\widehat{\boldsymbol{\mu}}(\tau)$. So we omit details for the corresponding theorems.

4. ON TIME-VARYING VARX

As a further application, we study a class of time-varying VARX(p,q) models of the form:

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

where $\boldsymbol{Z}_t = z_t \otimes \boldsymbol{I}_d$, $z_t = (1, \boldsymbol{y}_{t-1}^\top, \ldots, \boldsymbol{y}_{t-p}^\top, \boldsymbol{x}_{t-1}^\top, \ldots, \boldsymbol{x}_{t-q}^\top, \boldsymbol{x}_t^\top)^\top$, and $\boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau_t)\boldsymbol{\epsilon}_t$. Here, $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{d,t})^\top$ is a d-dimensional vector of endogenous variables, $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{m,t})^\top$ is an m-dimensional vector of exogenous variables, and both d and m are finite positive integers. We further allow $\boldsymbol{x}_t$ to follow a VMA(∞) process, see Assumption 5.3. Consequently, $\boldsymbol{y}_t$ itself also admits a VMA(∞) representation and thus we are able to establish a new estimation theory for the time-varying VARX models. Here, $\{\boldsymbol{A}_j(\tau)\}$ and $\{\boldsymbol{B}_j(\tau)\}$ are the $d \times d$ and $d \times m$ coefficient matrices. Also, $\boldsymbol{\omega}(\tau)$ is an unknown deterministic function that has full row rank uniformly in $\tau \in [0,1]$, and captures the heteroscedasticity over time. Obviously, we have

$$\boldsymbol{\beta}(\tau) = \mathrm{vec}(\boldsymbol{\mu}(\tau), \boldsymbol{A}(\tau), \boldsymbol{B}(\tau), \boldsymbol{B}_0(\tau)), \tag{4.2}$$

where $\boldsymbol{A}(\tau) = (\boldsymbol{A}_1(\tau), \ldots, \boldsymbol{A}_p(\tau))$ and $\boldsymbol{B}(\tau) = (\boldsymbol{B}_1(\tau), \ldots, \boldsymbol{B}_q(\tau))$.[1] In what follows, we focus on $\boldsymbol{\beta}(\tau)$.

4.1. Nonparametric Estimation

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

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

We estimate $\{\boldsymbol{\beta}(\tau), \boldsymbol{\beta}^{(1)}(\tau)\}$ using the kernel weighted least-square criterion:

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

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

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

[1] The specific form that $B_0(\cdot)$ is singled out in equation (4.2) allows us to rewrite the VARX(p,q) model in a VARX(1,0) form, which helps us to derive the asymptotic distributions of dynamic multipliers by using the Delta method (see Lütkepohl (2005, p. 403) for more details).

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

where $s = d + d^2p + (q+1)md$, $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_T^\top)^\top$, $\boldsymbol{K}_\tau = \text{diag}\{K_h(\tau_1 - \tau), \ldots, K_h(\tau_T - \tau)\} \otimes \boldsymbol{I}_d$, and

$$\boldsymbol{Z}_\tau = \begin{pmatrix} \boldsymbol{Z}_1^\top & \boldsymbol{Z}_1^\top \frac{\tau_1 - \tau}{h} \\ \vdots & \vdots \\ \boldsymbol{Z}_T^\top & \boldsymbol{Z}_T^\top \frac{\tau_T - \tau}{h} \end{pmatrix}.$$

The following assumptions are necessary for the theoretical development.

Assumption 5.

1. The roots of $\boldsymbol{I}_d - \boldsymbol{A}_1(\tau)L - \cdots - \boldsymbol{A}_p(\tau)L^p = \boldsymbol{0}_d$ all lie outside the unit circle uniformly in $\tau \in [0,1]$.
2. Each element of $\boldsymbol{\beta}(\tau)$ is second order continuously differentiable on $[0,1]$ and $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta}(0)$ for $\tau < 0$.
3. Suppose that $\boldsymbol{x}_t = \boldsymbol{g}(\tau_t) + \sum_{j=0}^{\infty} \boldsymbol{D}_j(\tau_t)\boldsymbol{v}_{t-j}$ for $t \geq 1$, and $\boldsymbol{x}_t = \boldsymbol{g}(0) + \sum_{j=0}^{\infty} \boldsymbol{D}_j(0)\boldsymbol{v}_{t-j}$ for $t \leq 0$, where $\boldsymbol{g}(\cdot)$ and $\boldsymbol{D}_j(\cdot)$ are $m \times 1$ and $m \times m$, respectively.
   Each component of $\boldsymbol{g}(\cdot)$ and $\boldsymbol{D}_j(\cdot)$ is second order continuously differentiable on $[0,1]$. For $\ell = 0, 1$, $\sup_{\tau \in [0,1]} \sum_{j=1}^{\infty} j \|\boldsymbol{D}_j^{(\ell)}(\tau)\| < \infty$, where $\boldsymbol{D}_j^{(\ell)}(\tau)$ denotes the ℓth derivative of $\boldsymbol{D}_j(\tau)$ and $\boldsymbol{D}_j^{(0)}(\tau) \equiv \boldsymbol{D}_j(\tau)$.
4. Each component of $\boldsymbol{\omega}(\tau)$ is second order continuously differentiable on $[0,1]$. Moreover, $\boldsymbol{\Omega}(\tau) = \boldsymbol{\omega}(\tau)\boldsymbol{\omega}(\tau)^\top$ is positive definite for $\forall \tau \in [0,1]$, and $\boldsymbol{\omega}(\tau) = \boldsymbol{\omega}(0)$ for $\tau < 0$.
5. Let $\boldsymbol{e}_t = (\boldsymbol{\epsilon}_t^\top, \boldsymbol{v}_{t+1}^\top)^\top$ and $\{\boldsymbol{e}_t\}_{t=-\infty}^{\infty}$ form a sequence of martingale differences such that $E(\boldsymbol{e}_t | \mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_t = \sigma\{\boldsymbol{e}_t, \boldsymbol{e}_{t-1}, \ldots\}$. Also, suppose that $E(\boldsymbol{e}_t\boldsymbol{e}_t^\top | \mathcal{F}_{t-1}) = \begin{pmatrix} \boldsymbol{I}_d & \boldsymbol{\rho}_{\epsilon v} \\ \boldsymbol{\rho}_{\epsilon v}^\top & \boldsymbol{I}_m \end{pmatrix}$ almost surely (a.s.), and $\sup_{t \geq 1} E\|\boldsymbol{e}_t\|^\delta < \infty$ for some $\delta > 4$.

Assumption 5.2 is standard in the literature (Li and Racine, 2007, p. 9), so the discussions are omitted. Assumption 5.5 is also standard and assumes that the innovation errors follow a martingale difference structure, which is identical to those used in Phillips and Lee (2013) for example.

We now comment on the rest of the conditions of Assumption 5. Assumption 5.1 ensures that $\boldsymbol{y}_t$ in model (4.1) is neither a unit-root process nor an explosive process, and can be regarded as an extension of those used for the classical multivariate dynamic models (e.g., Hamilton, 1994, p. 259). Note that Sun et al. (2021) assume that a time-varying univariate ARX process is β-mixing. We employ an alternative approach by utilizing the MA structure of ARX processes without imposing any mixing condition. Assumption 5.3 formulates a time-varying VMA(∞) process that nests many different processes as special cases as shown in Examples 1 and 2. Assumption 5.4 imposes certain heteroscedasticity.

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

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

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

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

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

2. *If, in addition, $\sup_{t\geq 1} E\left[\|\boldsymbol{e}_t\|^4 \,|\mathcal{F}_{t-1}\right] < \infty$ a.s. and $\frac{T^{1-4/\delta}h}{\log T} \to \infty$, then for $\forall \tau \in [0,1]$*

$$\widehat{\boldsymbol{V}}(\tau) \to_P \boldsymbol{V}(\tau),$$

where $\widehat{\boldsymbol{V}}(\tau) = \widehat{\boldsymbol{\Sigma}}_z^{-1}(\tau) \otimes \widehat{\boldsymbol{\Omega}}(\tau)$, $\widehat{\boldsymbol{\Sigma}}_z(\tau) = (\frac{1}{T}\sum_{t=1}^{T} K_h(\tau_t - \tau))^{-1}\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{z}_t\boldsymbol{z}_t^\top K_h(\tau_t - \tau)$, $\widehat{\boldsymbol{\Omega}}(\tau) = (\frac{1}{T}\sum_{t=1}^{T} K_h(\tau_t - \tau))^{-1}\frac{1}{T}\sum_{t=1}^{T} \widehat{\boldsymbol{\eta}}_t\widehat{\boldsymbol{\eta}}_t^\top K_h(\tau_t - \tau)$, and $\widehat{\boldsymbol{\eta}}_t = \boldsymbol{y}_t - \boldsymbol{Z}_t^\top\widehat{\boldsymbol{\beta}}(\tau_t)$.

The first result of Theorem 4.1 establishes the asymptotic distribution of $\widehat{\boldsymbol{\beta}}(\tau)$, while the second result ensures that the confidence interval can be constructed practically. Furthermore, using Theorem 4.1 and the Delta method, we can infer impulse responses (which are also known as dynamic multipliers) of $\boldsymbol{y}_t$ to $\boldsymbol{x}_t$. We present this result in Appendix B.1 of the Supplementary Material.

4.2. Semiparametric Estimation

In this subsection, we consider a semiparametric version of model (4.1) by allowing that some of $\boldsymbol{z}_t$ can be time-invariant. Let $\boldsymbol{z}_{1t}$ be a subset of $\boldsymbol{z}_t$, and let $\boldsymbol{z}_{2t}$ be a vector collecting the elements of $\boldsymbol{z}_t$ left out by $\boldsymbol{z}_{1t}$. Thus, (4.1) can be rewritten as

$$\boldsymbol{y}_t = \boldsymbol{Z}_{1t}^\top\boldsymbol{c} + \boldsymbol{Z}_{2t}^\top\boldsymbol{\theta}(\tau) + \boldsymbol{\eta}_t, \qquad \textbf{(4.6)}$$

where $\boldsymbol{Z}_{1t} = \boldsymbol{z}_{1t} \otimes \boldsymbol{I}_d$, $\boldsymbol{Z}_{2t} = \boldsymbol{z}_{2t} \otimes \boldsymbol{I}_d$, $\boldsymbol{c}$ and $\boldsymbol{\theta}(\tau)$ are the corresponding subsets of $\boldsymbol{\beta}(\tau)$ and are of dimensions s_1 and s_2, respectively. Equation (4.6) is a semiparametric model.

After profiling out the nonparametric part $\boldsymbol{\theta}(\tau)$, we obtain

$$\widetilde{\boldsymbol{y}}_t \simeq \widetilde{\boldsymbol{Z}}_{1t}^\top\boldsymbol{c} + \boldsymbol{\eta}_t,$$

where $\widetilde{\boldsymbol{y}}_t = \boldsymbol{y}_t - \boldsymbol{Z}_{2t}^\top\boldsymbol{s}(\tau_t)\boldsymbol{y}$, $\widetilde{\boldsymbol{Z}}_{1t} = \boldsymbol{Z}_{1t} - \boldsymbol{Z}_1^\top\boldsymbol{s}^\top(\tau_t)\boldsymbol{Z}_{2t}$, $\boldsymbol{s}(\tau) = (\boldsymbol{I}_{s_2}, \mathbf{0}_{s_2})(\boldsymbol{Z}_{2,\tau}^\top\boldsymbol{K}_\tau\boldsymbol{Z}_{2,\tau})^{-1}$ $\boldsymbol{Z}_{2,\tau}^\top\boldsymbol{K}_\tau$, $\boldsymbol{Z}_1 = (\boldsymbol{Z}_{11}, \ldots, \boldsymbol{Z}_{1T})^\top$ and $\boldsymbol{Z}_{2,\tau} = \begin{pmatrix} \boldsymbol{Z}_{21}^\top & \boldsymbol{Z}_{21}^\top\frac{\tau_1 - \tau}{h} \\ \vdots & \vdots \\ \boldsymbol{Z}_{2T}^\top & \boldsymbol{Z}_{2T}^\top\frac{\tau_T - \tau}{h} \end{pmatrix}$.

Since $\boldsymbol{\eta}_t$ has time-varying unconditional variance, we use the weighted least squares (WLS) method to estimate $\boldsymbol{c}$ in order to improve efficiency. The WLS

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

estimator of $\boldsymbol{c}$ is then given by

$$\widehat{\boldsymbol{c}}_{WLS} = \left(\sum_{t=1}^{T} \widetilde{\boldsymbol{Z}}_{1t} \widehat{\boldsymbol{\Omega}}^{-1}(\tau_t) \widetilde{\boldsymbol{Z}}_{1t}^{\top}\right)^{-1} \sum_{t=1}^{T} \widetilde{\boldsymbol{Z}}_{1t} \widehat{\boldsymbol{\Omega}}^{-1}(\tau_t) \widetilde{\boldsymbol{y}}_t, \tag{4.7}$$

where $\widehat{\boldsymbol{\Omega}}(\tau_t) = (\frac{1}{T}\sum_{s=1}^{T} K_h(\tau_s - \tau_t))^{-1} \frac{1}{T}\sum_{s=1}^{T} \widehat{\boldsymbol{\eta}}_s \widehat{\boldsymbol{\eta}}_s^{\top} K_h(\tau_s - \tau_t)$ for $t = 1, \ldots, T$, and $\widehat{\boldsymbol{\eta}}_t = \boldsymbol{y}_t - \boldsymbol{Z}_t^{\top} \widehat{\boldsymbol{\beta}}(\tau_t)$.

Finally, $\boldsymbol{\theta}(\tau)$ of (4.6) can be estimated by[2]

$$\widehat{\boldsymbol{\theta}}(\tau) = (\boldsymbol{I}_{s_2}, \boldsymbol{0}_{s_2})(\boldsymbol{Z}_{2,\tau}^{\top} \boldsymbol{K}_\tau \boldsymbol{Z}_{2,\tau})^{-1} \boldsymbol{Z}_{2,\tau}^{\top} \boldsymbol{K}_\tau (\boldsymbol{y} - \boldsymbol{Z}_1 \widehat{\boldsymbol{c}}_{WLS}). \tag{4.8}$$

We require the following assumption to facilitate the development.

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

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

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

THEOREM 4.2. *Let Assumptions 4–6 hold.*

1. *For (4.7), we have as* $T \to \infty$,

$$\sqrt{T}(\widehat{\boldsymbol{c}}_{WLS} - \boldsymbol{c}) \to_D N\left(\boldsymbol{0}, \boldsymbol{\Delta}_{\boldsymbol{c}}^{-1}\right),$$

where $\boldsymbol{\Delta}_{\boldsymbol{c}} = \int_0^1 (\boldsymbol{\Sigma}_{z_{1,1}}(\tau) - \boldsymbol{\Sigma}_{z_{1,2}}(\tau)\boldsymbol{\Sigma}_{z_{2,2}}^{-1}(\tau)\boldsymbol{\Sigma}_{z_{1,2}}^{\top}(\tau)) \otimes \boldsymbol{\Omega}^{-1}(\tau)\mathrm{d}\tau$, *in which for* $j_1, j_2 \in \{1,2\}$, $\boldsymbol{\Sigma}_{z_{j_1,j_2}}(\tau) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(z_{j_1 t} z_{j_2 t}^{\top}) K_h(\tau_t - \tau)$.

2. *For (4.8), we also have for any given* $\tau \in (0,1)$ *and as* $T \to \infty$,

$$\sqrt{Th}\left(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau) - \frac{1}{2}h^2 \tilde{c}_2 \boldsymbol{\theta}^{(2)}(\tau) + o_P(h^2)\right) \to_D N(\boldsymbol{0}, \tilde{v}_0 \boldsymbol{\Delta}_{\boldsymbol{\theta}}(\tau)),$$

where $\boldsymbol{\Delta}_{\boldsymbol{\theta}}(\tau) = \boldsymbol{\Sigma}_{z_2}^{-1}(\tau) \otimes \boldsymbol{\Omega}(\tau)$.

Similar to Theorem 4.1, both $\boldsymbol{\Delta}_{\boldsymbol{c}}$ and $\boldsymbol{\Delta}_{\boldsymbol{\theta}}(\tau)$ can be easily estimated by replacing the unknown quantities with their estimators.

In the following section, we conduct numerical studies to evaluate the proposed estimation methods and their theoretical properties.

[2]The kernel-weighted method uses local observations around time τ to estimate $\boldsymbol{\theta}(\tau)$, so the variance of $\boldsymbol{\eta}_t$ is automatically approximated by a constant due to the nature of the nonparametric kernel method. As a sequence, there is no need to correct the unconditional heteroscedasticity when estimating $\boldsymbol{\theta}(\tau)$.

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

5. NUMERICAL STUDIES

We present a computational implementation procedure in Section 5.1, and then conduct extensive numerical studies in Sections 5.2 and 5.3.

5.1. Computational Implementation

Throughout the numerical studies, the Epanechnikov kernel $K(u) = 0.75(1-u^2)I(|u| \le 1)$ is adopted. When studying the trending term of the VMA(∞) model, we use the "leave-($2k$+1)-out" version of cross-validation method of Chu and Marron (1991), as the error components are serially correlated. Specifically,

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

where $\widehat{\boldsymbol{\mu}}_{k,h}(\tau) = \left[\sum_{t:|t-\tau T|>k} K\left(\frac{\tau_t-\tau}{h}\right)\right]^{-1} \sum_{t:|t-\tau T|>k} \boldsymbol{x}_t K\left(\frac{\tau_t-\tau}{h}\right)$ and $k=5$. For the DWB method, we follow the suggestions of Bühlmann (1998), Shao (2010), and Friedrich et al. (2020) by choosing $\widetilde{h} = c_0 \cdot \widehat{h}_{mcv}^{5/9}$ with $c_0 = 2$, $a(x) = \frac{\int_{-1}^{1} w(u)w(u+|x|)\mathrm{d}u}{\int_{-1}^{1} w^2(u)\mathrm{d}u}$ with $w(u) = \frac{u}{0.43} I(u \in [0, 0.43)) + I(u \in [0.43, 0.57]) + \frac{1-u}{0.43} I(u \in (0.57, 1])$, and $l = 1.75 \cdot (T\widehat{h}_{mcv})^{1/3}$.

When studying the time-varying VARX model, we use the leave-one-out method, since the error terms of VARX model are mutually uncorrelated:

$$\widehat{h}_{cv} = \arg\min_h \sum_{t=1}^{T} \left(\boldsymbol{y}_t - \boldsymbol{Z}_{1t}^\top \widehat{\boldsymbol{c}}_{h,-t} - \boldsymbol{Z}_{2t}^\top \widehat{\boldsymbol{\theta}}_{h,-t}(\tau_t)\right)^\top \left(\boldsymbol{y}_t - \boldsymbol{Z}_{1t}^\top \widehat{\boldsymbol{c}}_{h,-t} - \boldsymbol{Z}_{2t}^\top \widehat{\boldsymbol{\theta}}_{h,-t}(\tau_t)\right), \tag{5.2}$$

where $\widehat{\boldsymbol{c}}_{h,-t}$ and $\widehat{\boldsymbol{\theta}}_{h,-t}(\cdot)$ are obtained using (4.7) and (4.8) but leaving the tth observation out. For the above minimization, we set a predetermined sequence of h's from a wide range, say from 0.1 to 0.6 with an increment 0.02.

5.2. Simulation Results

We first evaluate the finite sample performance of the DWB procedure presented in Section 3. Consider a multivariate time series data generating process (DGP):

$$\boldsymbol{x}_t = \boldsymbol{\mu}(\tau_t) + \boldsymbol{e}_t, \quad \boldsymbol{e}_t = \boldsymbol{A}(\tau_t)\boldsymbol{e}_{t-1} + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, T, \tag{5.3}$$

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

$$\boldsymbol{A}(\tau) = \begin{bmatrix} 0.1d\exp(-0.5+\tau) & (\tau-0.5)^3 \\ (\tau-0.5)^3 & 0.1d + 0.3\sin(\pi\tau) \end{bmatrix}$$

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

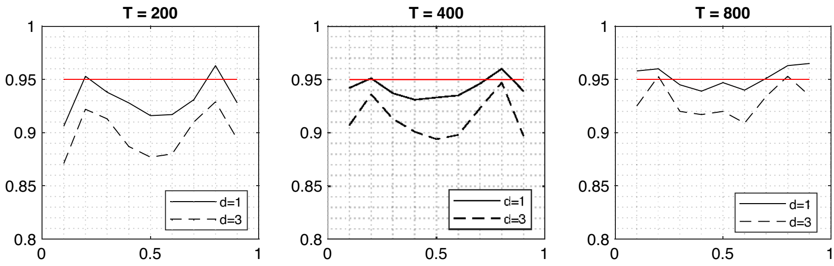

FIGURE 1. Point-wise coverage probabilities for $\boldsymbol{\mu}(\cdot)$.

with $d \in \{1,3\}$, which corresponds to low and moderate persistence in the dynamics of innovations, respectively. In addition, let the sample size be $T \in \{200, 400, 800\}$ and conduct 1,000 replications for each choice of T.

In order to evaluate the finite sample performance, we calculate the point-wise coverage rate associated with $\boldsymbol{\mu}(\cdot)$ based on the DWB procedure with $J = 1{,}000$ bootstrap replications. Specifically, we consider the coverage at $\tau = 0.1, \ldots, 0.9$, and use the nominal coverage 95%. For each given τ, the coverage probability is first calculated for each component of $\boldsymbol{\mu}(\cdot)$. We then take the average across these elements.

The coverage rates are plotted in Figure 1. A few facts emerge. First, for the small positive correlation case, the coverage probabilities are close to the nominal level even when the sample size is relatively small (i.e., $T = 200$). Second, the finite sample coverage probabilities decrease with the increase of d (which measures the extent of serial correlations in error innovations) for all the sample sizes considered. Intuitively, the data may deviate from the trend functions in clusters in the case of strong positive correlation, which causes the nonparametric estimate to go through these clusters and thus, to deviate significantly from the true trend. Interestingly, all time points are similarly affected. Third, for the strong positive correlation case, the finite sample coverage probabilities are smaller than their nominal level (95%) for small T, but are fairly close to 95% as T increases.

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

$$\begin{aligned}
&\text{DGP 1: } \boldsymbol{y}_t = \boldsymbol{\mu}(\tau_t) + \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \boldsymbol{B}_1(\tau_t) x_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau_t)\boldsymbol{\epsilon}_t,\\
&\text{DGP 2: } \boldsymbol{y}_t = \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \boldsymbol{B}_1(\tau_t) x_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau_t)\boldsymbol{\epsilon}_t, \qquad \textbf{(5.4)}
\end{aligned}$$

where $\boldsymbol{\epsilon}_t$'s are i.i.d. draws from $N(\boldsymbol{0}_{2\times 1}, \boldsymbol{I}_2)$, $\boldsymbol{\mu}(\tau) = [\sin(\pi\tau), \cos(\pi\tau)]^\top$, $x_t = 0.4x_{t-1} + v_t$ with $v_t \sim N(0,1)$, and

$$\boldsymbol{A}_1 = \begin{bmatrix} 0.1 + 0.3d & -0.1 \\ -0.1 & 0.1 + 0.3d \end{bmatrix},$$

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

Table 1. Empirical coverage probabilities for $\boldsymbol{\mu}(\tau)$, $\boldsymbol{A}_1$ and $\boldsymbol{B}_1(\cdot)$.

		$d=1$			$d=3$		
	T	$\boldsymbol{\mu}(\tau)$	$\boldsymbol{A}_1$	$\boldsymbol{B}_1(\cdot)$	$\boldsymbol{\mu}(\tau)$	$\boldsymbol{A}_1$	$\boldsymbol{B}_1(\cdot)$
	200	0.893	0.903	0.911	0.853	0.873	0.908
DGP 1	400	0.911	0.929	0.931	0.893	0.914	0.930
	800	0.924	0.937	0.939	0.907	0.925	0.936
	200	—	0.932	0.922	—	0.924	0.921
DGP 2	400	—	0.945	0.936	—	0.942	0.936
	800	—	0.949	0.939	—	0.945	0.938

Table 2. The ratios of the RMSEs of OLS estimator relative to that of WLS (the levels of RMSE are reported for WLS in brackets).

		$d=1$		$d=3$	
	T	OLS	WLS	OLS	WLS
	200	1.061	[0.144]	1.061	[0.142]
DGP 1	400	1.077	[0.097]	1.079	[0.091]
	800	1.082	[0.067]	1.083	[0.062]
	200	1.062	[0.135]	1.062	[0.124]
DGP 2	400	1.078	[0.093]	1.080	[0.083]
	800	1.082	[0.065]	1.081	[0.058]

$$\boldsymbol{\omega}(\tau) = \begin{bmatrix} 1.5+0.2\exp(0.5-\tau) & 0 \\ 0.1\exp(0.5-\tau) & 1.5+0.5(\tau-0.5)^2 \end{bmatrix}.$$

Similarly, we set $d \in \{1,3\}$, which corresponds to low and moderate persistence in the VAR dynamics. The coverage rates are calculated in a similar manner as above, so we omit the details.

As shown in Table 1, the coverage rates move toward 95% as the sample size goes up. In addition, similar to the trending estimation case, the rates are worse with a larger value of d, which is not surprising.

Finally, we compare the performance of WLS and OLS estimators of $\boldsymbol{A}_1$. Here, the WLS estimator is given in (4.7), while the OLS estimator is defined as

$$\widehat{\boldsymbol{c}}_{OLS} = \left(\sum_{t=1}^{T} \widetilde{\boldsymbol{Z}}_{1t}\widetilde{\boldsymbol{Z}}_{1t}^{\top}\right)^{-1} \sum_{t=1}^{T} \widetilde{\boldsymbol{Z}}_{1t}\widetilde{\boldsymbol{y}}_t.$$

Table 2 reports the ratios of the root mean squared errors (RMSEs) of the OLS estimator relative to those of the WLS estimator. The levels (rather than the ratios) of RMSEs are reported for WLS in brackets. Clearly, OLS is inefficient. In

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

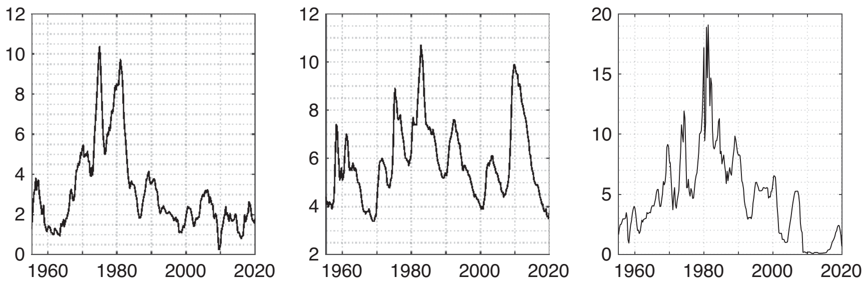

FIGURE 2. Plots of the inflation (left), the unemployment rate (middle), and the interest rate (right).

addition, the performance of WLS relative to OLS is improved with the increased sample size. Moreover, our simulation results also show that the improvement of the WLS procedure relative to OLS is insensitive to the persistence of VAR dynamics, as well as the presence of time-varying intercepts.

5.3. A Real Data Example

In the empirical study, we infer the trend value of inflation (i.e., trend inflation, see Stock and Watson, 2016a for more details) and the trend value of unemployment rate (i.e., the natural rate of unemployment, NAIRU, see Staiger, Stock, and Watson, 1997 for more details). The trend inflation and NAIRU are centrally positioned in setting monetary policy since the Federal Reserve Bank aims to mitigate deviations of inflation and unemployment from their long-run targets (Primiceri, 2006; Stock and Watson, 2016a). The estimation is conducted in exactly the same way as in Section 5.1, so we will not repeat the details unless necessary.

Specifically, we estimate the time-varying VMA(∞) model (3.1) using three commonly adopted macroeconomic variables of the literature (Primiceri, 2005; Cogley, Primiceri, and Sargent, 2010); the inflation rate (measured by the 100 times the year-over-year log change in the GDP deflator), the unemployment rate, and the interest rate (measured by the average value for the Federal funds rates over the quarter). Although we are not interested in the trend of interest rates, we include this variable within the system in order to capture more dynamics and be consistent with the literature. The data are quarterly observations measured at an annual rate from 1954:Q3 to 2020:Q1, and are collected from the Federal Reserve Bank of St. Louis economic database. Figure 2 plots the three variables.

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

$$\boldsymbol{\mu}_t = \lim_{p\to\infty} E_t(\boldsymbol{x}_{t+p}) = (\boldsymbol{I}_2 - \boldsymbol{A}_{1t} - \boldsymbol{A}_{2t})^{-1}\boldsymbol{a}_t, \tag{5.5}$$

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

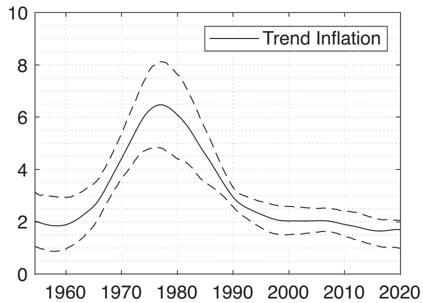

FIGURE 3. The estimated trends (i.e., $\boldsymbol{\mu}(\cdot)$) of inflation and unemployment as well as the associated 95% bootstrap confidence intervals.

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

Figure 3 plots the estimates of the trend inflation and the NAIRU (i.e., $\widehat{\boldsymbol{\mu}}(\tau)$), as well as the 95% bootstrap confidence intervals. Here, we use the bootstrap method to construct confidence intervals, which remains robust whether the VMA coefficients are time-varying or not. First, we can see that the length of confidence intervals change over time. For example, the estimation uncertainty was high in the 1980s, but this uncertainty decreases dramatically after this period. This fact implies that our time varying VMA model is much better than the constant VMA model. Note that constant VMA coefficients correspond to a constant asymptotic variance over $\tau \in [0, 1]$ according to Theorem 3.1. In addition, it is obvious that the underlying trend of inflation was high in the 1970s, but decreased in the subsequent period. After the Great Moderation, the long-run level of inflation was below, but quite close to the Federal Reserve's target of 2%, indicating that inflation is more anchored now than it was in the 1970s. However, the NAIRU is less persistent and fluctuates over time. In particular, the NAIRU increased rapidly during the "Second Oil Crisis" and "Global Financial Crisis."

To provide more empirical evidence on the time-varying VMA coefficients, we also estimate the time-varying volatility of $\boldsymbol{x}_t$, that is, the square root of diagonal elements of

$$E(\boldsymbol{e}_t\boldsymbol{e}_t^\top) = \sum_{j=0}^{\infty} \mathbf{B}_j(\tau)\mathbf{B}_j^\top(\tau).$$

Figure 4 plots the estimated volatilities of both the inflation and unemployment rate, clearly showing that their volatilities change over time, which further underlines the empirical importance of the proposed time-varying VMA model. We can also see that a decline in unconditional volatilities of exogenous shocks after 1980

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

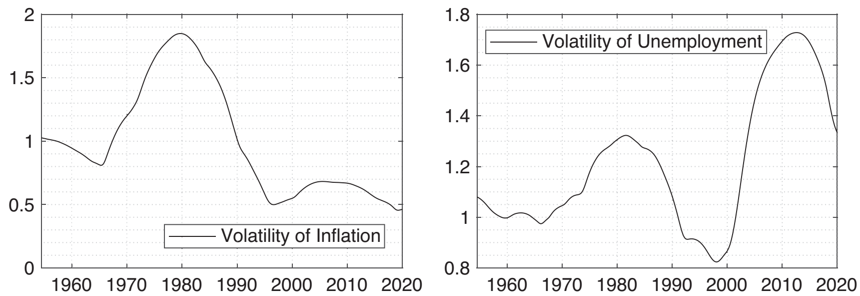

FIGURE 4. Estimates of the time-varying volatility of inflation (left panel) and unemployment rate (right panel).

in every subplot of Figure 4. Thus, our results support the explanation of "bad luck" (e.g., Primiceri, 2005; Sims and Zha, 2006).

6. CONCLUSION

In this paper, we introduce a class of time-varying VMA(∞) processes, and derive a set of asymptotic properties accordingly. Our investigation starts with decomposing the weighted sum of time-varying VMA(∞) processes into the long-run and transitory elements, known as the BN decomposition (Beveridge and Nelson, 1981; Phillips and Solo, 1992). As the long-run component of the decomposition yields a martingale approximation, it ensures the feasibility of achieving a variety of asymptotic properties for the multivariate case, for example, the law of large numbers, the uniform convergence, the central limit theorem, the bootstrap consistency, and the long-run covariance matrix estimation.

Furthermore, we show that these results can be readily applied when establishing inferences for many other dynamic time-varying models. In the empirical study, we apply the newly proposed framework to study the long-run level of inflation and the natural rate of unemployment. We find that (1) the long-run level of inflation is more anchored now and is close to the Federal Reserve's target of 2% after the beginning of the Great Moderation period, and (2) the natural rate of unemployment is less persistent and increases rapidly during the "Second Oil Crisis" and "Global Financial Crisis."

APPENDIX

In this Appendix, we provide some selected proofs of the main results. In what follows, M and $O(1)$ always stand for bounded constants, and may be different at each appearance.

Proof of Lemma 2.1. By the BN decomposition in Lemma B.3 in Appendix B of the Supplementary Material, we have

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

By using a functional central limit theorem for martingale difference sequences, for example, Theorem 18.2 in Billingsley (2013), we have $L_{T,1}\to_D\int_0^r\mathbb{B}(u)\mathrm{d}\boldsymbol{W}(u)$. Note that the condition $\limsup_{T\to\infty}\max_{1\le t\le T}\|\mathbb{B}_t(1)-\mathbb{B}(t/T)\|=0$ (plus the continuity condition on $\mathbb{B}(\cdot)$) is enough to ensure the convergence of conditional variance, that is, $\frac{1}{T}\sum_{t=1}^{\lfloor Tr\rfloor}\mathbb{B}_t(1)\mathbb{B}_t^\top(1)\to\int_0^r\mathbb{B}(u)\mathbb{B}^\top(u)\mathrm{d}u$. In addition, we have $L_{T,2}\to_P 0$ uniformly over $r\in[0,1]$ because of $E\left\|\widetilde{\mathbb{B}}_1(L)\boldsymbol{\epsilon}_0\right\|<\infty$ by Lemma B.3 in Appendix B of the Supplementary Material.

For $L_{T,3}$, we need to show that

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

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

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

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

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

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

The proof is now completed. □

Proof of Lemma 2.2. By Lemma B.3 in Appendix B of the Supplementary Material, we have $\boldsymbol{x}_t=\boldsymbol{\mu}_t+\mathbb{B}_t(1)\boldsymbol{\epsilon}_t+\widetilde{\mathbb{B}}_t(L)\boldsymbol{\epsilon}_{t-1}-\widetilde{\mathbb{B}}_t(L)\boldsymbol{\epsilon}_t$, which yields

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

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

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

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

Hence, $\|\boldsymbol{I}_{T,1}\| = O_P(\sqrt{d_T})$. Also, $\|\boldsymbol{I}_{T,2}\| = O_P(d_T)$ and $\|\boldsymbol{I}_{T,3}\| = O_P(d_T)$, since $\max_{t\geq 1}\left\|\boldsymbol{W}_{T,t}\right\| = O(d_T)$, $E\|\widetilde{\mathbb{B}}_1(L)\boldsymbol{\epsilon}_0\| < \infty$ and $E\|\widetilde{\mathbb{B}}_T(L)\boldsymbol{\epsilon}_t\| < \infty$ by Lemma B.3 in Appendix B of the Supplementary Material.

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

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

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

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

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

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

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

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

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

which yields

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

Consequently, we obtain

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

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

For $I_{T,6}$, by Lemma B.3 in Appendix B of the Supplementary Material, write

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

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

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

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

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

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

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

For $I_{T,7}$, using Lemma B.3 in Appendix B of the Supplementary Material, we have

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

For $I_{T,71}$, by Lemma B.3 in Appendix B of the Supplementary Material, we further write

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

In addition, similar to the proofs of $\boldsymbol{I}_{T,2}$ to $\boldsymbol{I}_{T,4}$, we can show that $I_{T,72}$ to $I_{T,74}$ are $O_P(d_T)$. Combining the above results, we have proved the case of $p = 0$. Similarly to the development of $p = 0$, we can complete the proof for the case of $p \geq 1$ given p is a fixed number. The details are omitted due to similarity. The proof is now completed. □

Proof of Lemma 2.3.
(1). By Lemma B.3 in Appendix B of the Supplementary Material, we have $\boldsymbol{x}_t = \boldsymbol{\mu}_t + \mathbb{B}_t(1)\boldsymbol{\epsilon}_t + \widetilde{\mathbb{B}}_t(L)\boldsymbol{\epsilon}_{t-1} - \widetilde{\mathbb{B}}_t(L)\boldsymbol{\epsilon}_t$.

We are then able to write

$$\begin{aligned}
&\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}\boldsymbol{W}_{T,t}(\tau)(\boldsymbol{x}_t-E(\boldsymbol{x}_t))\right\| \\
&\leq \sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}\boldsymbol{W}_{T,t}(\tau)\mathbb{B}_t(1)\boldsymbol{\epsilon}_t\right\| + \sup_{\tau\in[a,b]}\left\|\boldsymbol{W}_{T,1}(\tau)\widetilde{\mathbb{B}}_1(L)\boldsymbol{\epsilon}_0\right\| + \sup_{\tau\in[a,b]}\left\|\boldsymbol{W}_{T,T}(\tau)\widetilde{\mathbb{B}}_T(L)\boldsymbol{\epsilon}_T\right\| \\
&\quad + \sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T-1}\left(\boldsymbol{W}_{T,t+1}(\tau)\widetilde{\mathbb{B}}_{t+1}(L)-\boldsymbol{W}_{T,t}(\tau)\widetilde{\mathbb{B}}_t(L)\right)\boldsymbol{\epsilon}_t\right\| := I_{T,81}+I_{T,82}+I_{T,83}+I_{T,84},
\end{aligned}$$

where the definitions of $I_{T,8j}$ for $j=1,\ldots,4$ are obvious.

By Lemma B.4 in Appendix B of the Supplementary Material, we have $I_{T,81}=O_P\left(\sqrt{d_T\log T}\right)$. It is also easy to see that $I_{T,82}=O_P(d_T)$ and $I_{T,83}=O_P(d_T)$, 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 in Appendix B of the Supplementary Material. Thus, we need only to consider $I_{T,84}$ 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 in Appendix B of the Supplementary Material;

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

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

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

Hence, write

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

The first result then follows.

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

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

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

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

$$+2\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(\tau))\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}(\boldsymbol{B}_{j+r,t}\otimes\boldsymbol{B}_{j,t})\mathrm{vec}\left(\boldsymbol{\epsilon}_{t-j}\epsilon_{t-j-r}\right)\right\|:=2I_{T,91}+I_{T,92}+2I_{T,93},$$

wherein $I_{T,91}=O_P(\sqrt{d_T\log T})$ by a proof similar to the first result of this lemma.

Consider $I_{T,92}$. Using Lemma B.3 in Appendix B of the Supplementary Material, write

$$I_{T,92}\leq\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes\boldsymbol{W}_{T,t}(\tau))\mathbb{B}_t^0(1)\left(\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\right)-\mathrm{vec}\,(\boldsymbol{I}_d)\right)\right\|$$

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

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

$$:=I_{T,101}+I_{T,102}+I_{T,103}+I_{T,104}.$$

By Lemma B.5 in Appendix B of the Supplementary Material, we have $I_{T,101}=O_P\left(\sqrt{d_T\log T}\right)$. Also, $I_{T,102}=O_P(d_T)$ and $I_{T,103}=O_P(d_T)$, because $\|\widetilde{\mathbb{B}}_1^0(1)\|<\infty$ and $\|\widetilde{\mathbb{B}}_T^0(1)\|<\infty$ by Lemma B.3 in Appendix B of the Supplementary Material. Similar to the proof of the first result, for $I_{T,24}$, we write

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

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

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

where we have used the following facts:

$$(1).\ T^{4/\delta}d_T\log T\to 0;\ (2).\ \max_{t\geq 1}\left\|\widetilde{\mathbb{B}}_t^0(L)\mathrm{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\right)\right\|=O_P(T^{2/\delta});$$

$$(3).\ \sup_{\tau\in[a,b]}\sum_{t=1}^{T-1}\|\boldsymbol{W}_{T,t+1}(\tau)-\boldsymbol{W}_{T,t}(\tau)\|=O(d_T);\ (4).\ \sum_{t=1}^{T-1}\left\|\widetilde{\mathbb{B}}_{t+1}^0(1)-\widetilde{\mathbb{B}}_t^0(1)\right\|=O(1).$$

Then we can conclude that $I_{T,104}=O_P\left(\sqrt{d_T\log T}\right)$.

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

We now consider $I_{T,93}$. Using Lemma B.3 in Appendix B of the Supplementary Material, we have

$$
\begin{aligned}
&\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t}(\tau))\sum_{r=1}^{\infty}\sum_{j=0}^{\infty}(\boldsymbol{B}_{j+r,t}\otimes \boldsymbol{B}_{j,t})\text{vec}\left(\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}_{t-j-r}\right)\right\|\\
\leq\ &\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T}(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t}(\tau))\boldsymbol{\zeta}_t\boldsymbol{\epsilon}_t\right\|+\sup_{\tau\in[a,b]}\left\|(\boldsymbol{I}_d\otimes W_{T,1}(\tau))\sum_{r=1}^{\infty}\widetilde{\mathbb{B}}_1^r(L)\text{vec}\left(\boldsymbol{\epsilon}_0\boldsymbol{\epsilon}_{-r}^{\top}\right)\right\|\\
&+\sup_{\tau\in[a,b]}\left\|(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,T}(\tau))\sum_{r=1}^{\infty}\widetilde{\mathbb{B}}_T^r(L)\text{vec}\left(\boldsymbol{\epsilon}_T\boldsymbol{\epsilon}_{T-r}^{\top}\right)\right\|\\
&+\sup_{\tau\in[a,b]}\left\|\sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\left((\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t+1}(\tau))\widetilde{\mathbb{B}}_{t+1}^r(L)-(\boldsymbol{I}_d\otimes \boldsymbol{W}_{T,t}(\tau))\widetilde{\mathbb{B}}_t^r(L)\right)\text{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^{\top}\right)\right\|\\
:=\ &I_{T,111}+I_{T,112}+I_{T,113}+I_{T,114},
\end{aligned}
$$

where $\boldsymbol{\zeta}_t$ is defined in Lemma B.5 in Appendix B of the Supplementary Material.

By Lemma B.5 in Appendix B of the Supplementary Material, $I_{T,111}=O_P\left(\sqrt{d_T\log T}\right)$. Moreover, $I_{T,112}=O_P(d_T)$ and $I_{T,113}=O_P(d_T)$, 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 in Appendix B of the Supplementary Material. For $I_{T,114}$, we write

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

where we have used the following results:

$$
(1).\ T^{4/\delta}d_T\log T\to 0;\ (2).\ \max_t\left\|\sum_{r=1}^{\infty}\widetilde{\mathbb{B}}_t^r(L)\text{vec}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-r}^{\top}\right)\right\|=O_P(T^{2/\delta});
$$

$$
(3).\ \sum_{t=1}^{T-1}\sum_{r=1}^{\infty}\|\widetilde{\mathbb{B}}_{t+1}^r(1)-\widetilde{\mathbb{B}}_t^r(1)\|=O(1);\ (4).\ \sup_{\tau\in[a,b]}\sum_{t=1}^{T-1}\left\|\boldsymbol{W}_{T,t+1}(\tau)-\boldsymbol{W}_{T,t}(\tau)\right\|=O(d_T).
$$

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

Proof of Lemma 2.4. Similar to the proof of Lemma 2.2, we have

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

as $d_T=o(1)$.

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

Since

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

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

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

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

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

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

since $\sum_{t=1}^{T}\left\|\boldsymbol{W}_{T,t}\right\|=O(1)$ and $\max_{t\ge1}\left\|\boldsymbol{W}_{T,t}\right\|=O(d_T)$. By Lemma B.1 in Appendix B of the Supplementary Material, the proof is now completed. □

Proof of Lemma 2.5. By $d_T^{-1/2}\sum_{t=1}^{T}\boldsymbol{W}_{T,t}(\boldsymbol{x}_t-E(\boldsymbol{x}_t))\to_D N(\boldsymbol{0},\boldsymbol{\Sigma}_{\boldsymbol{W}})$ and using a multivariate version of Polya's theorem (p. 23, Bhattacharya and Rao, 1986), we have

$$\sup_{\boldsymbol{w}\in\mathbb{R}^d}\left|\Pr\left(d_T^{-1/2}\sum_{t=1}^{T}\boldsymbol{W}_{T,t}(\boldsymbol{x}_t-E(\boldsymbol{x}_t))\le\boldsymbol{w}\right)-\Phi(\boldsymbol{w};\boldsymbol{\Sigma}_{\boldsymbol{W}})\right|=o(1),$$

where $\Phi(\cdot;\boldsymbol{\Sigma}_{\boldsymbol{W}})$ denotes the CDF function of multivariate normal variables with zero mean and variance $\boldsymbol{\Sigma}_{\boldsymbol{W}}$. Hence, it is enough to obtain $d_T^{-1/2}\sum_{t=1}^{T}\boldsymbol{W}_{T,t}(\boldsymbol{x}_t-E(\boldsymbol{x}_t))\xi_t^*\to_{D^*}N(\boldsymbol{0},\boldsymbol{\Sigma}_{\boldsymbol{W}})$ in order to show

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

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

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

In the following step, we first show that

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

and then prove its normality by blocking techniques.

Conditioning on the original sample, we have

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

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

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

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

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

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

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

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

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

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

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

by the proof of Lemma 2.2.

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

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

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

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

Conditioning on the original sample, we now employ standard arguments for using a block technique to show the asymptotic normality. Now, let $Z_T^*(\tau)=\sum_{j=1}^k X_{T,j}^*(\tau)+\sum_{j=1}^k Y_{T,j}^*(\tau)$, where $X_{T,j}^*(\tau)=\frac{1}{\sqrt{d_T}}\sum_{t=B_j+1}^{B_j+r_1}\boldsymbol{d}^\top\boldsymbol{W}_{T,t}\boldsymbol{e}_t\xi_t^*$ and $Y_{T,j}^*(\tau)=\frac{1}{\sqrt{d_T}}\sum_{t=B_j+r_1+1}^{B_j+r_1+r_2}\boldsymbol{d}^\top\boldsymbol{W}_{T,t}\boldsymbol{e}_t\xi_t^*$, in which $B_j=(j-1)(r_1+r_2)$ and $k=\lceil T/(r_1+r_2)\rceil$.

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

$$\begin{aligned}
&EE^*\left(\sum_{j=1}^k Y_{T,j}^*(\tau)\right)^2=E\left(\sum_{j=1}^k E^*(Y_{T,j}^*(\tau))^2\right)\\
\leq\ &\frac{1}{d_T}\sum_{i=-r_2+1}^{r_2-1}a(i/l)\max_t\left\|E(\boldsymbol{e}_t\boldsymbol{e}_{t+i}^\top)\right\|\sum_{j=1}^k\sum_{t=B_j+r_1+1}^{B_j+r_1+r_2-i}\|\boldsymbol{W}_{T,t}\|\cdot\|\boldsymbol{W}_{T,t+i}\|\\
\leq\ &M\frac{1}{d_T}\max_{0\leq i\leq r_2-1}\sum_{j=1}^k\sum_{t=B_j+r_1+1}^{B_j+r_1+r_2-i}\|\boldsymbol{W}_{T,t}\|\cdot\|\boldsymbol{W}_{T,t+i}\|\leq\ Mkr_2d_T=o(1).
\end{aligned}$$

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

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

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

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

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

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

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

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

Combining the above results, $Z_T^* \to_{D^*} N\left(\boldsymbol{0}, \boldsymbol{d}^\top \boldsymbol{\Sigma}_{\boldsymbol{W}} \boldsymbol{d}\right)$. The proof is now completed. □

Proof of Lemma 2.6. Define $\boldsymbol{\Xi}_i = \frac{1}{d_T}\sum_{t=1}^{T-i} \boldsymbol{W}_{T,t} E\left(\boldsymbol{e}_t \boldsymbol{e}_{t+i}^\top\right) \boldsymbol{W}_{T,t+i}^\top$ with $\boldsymbol{e}_t = \sum_{j=0}^{\infty} \boldsymbol{B}_{j,t}\boldsymbol{\epsilon}_t$. Write

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

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

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

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

Hence, we have $I_{T,1} \to \boldsymbol{\Sigma}_{\boldsymbol{W}}$.

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

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

The proof is now completed. □

https://doi.org/10.1017/S0266466623000397 Published online by Cambridge University Press

SUPPLEMENTARY MATERIAL

Yan, Y, Gao, J., and Peng, B. (2023). Supplement to "Asymptotics for Time-Varying Vector MA(∞) Processes," *Econometric Theory* Supplementary Material. To view, please visit https://doi.org/10.1017/S0266466623000397.

References

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3), 817–858.

Beveridge, S., & 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.

Bhattacharya, R. N., & Rao, R. R. (1986). *Normal approximation and asymptotic expansions*. SIAM.

Billingsley, P. (2013). *Convergence of probability measures*. John Wiley & Sons.

Brüggemann, R., Jentsch, C., & Trenkler, C. (2016). Inference in VARs with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, 191(1), 69–85.

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

Casini, A. (2023). Theory of evolutionary spectra for heteroskedasticity and autocorrelation robust inference in possibly misspecified and nonstationary models. *Journal of Econometrics*, 235(2), 372–392.

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

Cogley, T., Primiceri, G. E., & Sargent, T. J. (2010). Inflation-gap persistence in the U.S. *American Economic Journal: Macroeconomics*, 2(1), 43–69.

Dahlhaus, R. (2012). Locally stationary processes. In T. Subba Rao, S. Subba Rao & C.R. Rao (eds.), *Handbook of statistics*, vol. 30 (pp. 351–413). Elsevier.

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

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

Doukhan, P. (2012). *Mixing: Properties and examples*, vol. 85. Springer Science & Business Media.

Fan, J., & Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, 11(6), 1031–1057.

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

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

Giraitis, L., Kapetanios, G., & Yates, T. (2014). Inference on stochastic time-varying coefficient models. *Journal of Econometrics*, 179(1), 46–65.

Hamilton, J. D. (1994). *Time series analysis*, vol. 2. Princeton University Press.

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

Hansen, B. E. (2001). The new econometrics of structural change: dating breaks in U.S. 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.

Kiefer, N. M., & Vogelsang, T. J. (2002). Heteroskedasticity-autocorrelation robust standard errors using the bartlett kernel without truncation. *Econometrica*, 70(5), 2093–2095.

Lenza, M., & Primiceri, G. E. (2022). How to estimate a vector autoregression after March 2020. *Journal of Applied Econometrics*, 37(4), 688–699.

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

Li, Q., & Racine, J. (2007). *Nonparametric econometrics theory and practice*. Princeton University Press.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

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

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

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

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

Phillips, P. C. B., & 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.

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

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

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

Sims, C., & Zha, T. (2006). Were there regime switches in U.S. monetary policy? *American Economic Review*, 96(1), 54–81.

Staiger, D. O., Stock, J. H., & Watson, M. W. (1997). How precise are estimates of the natural rate of unemployment? In C. D. Romer & D. H. Romer (eds.), *Reducing inflation: Motivation and strategy* (pp. 195–246). University of Chicago Press.

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

Stock, J. H., & Watson, M. W. (2016b). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In J. B. Taylor & H. Uhlig (eds.), *Handbook of macroeconomics*, vol. 2 (415–525). Elsevier.

Sun, Y., Hong, Y., Lee, T.-H., Wang, S., & Zhang, X. (2021). Time-varying model averaging. *Journal of Econometrics*, 222(2), 974–992.

Sun, Y., Phillips, P. C., & Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing. *Econometrica*, 76(1), 175–194.

Withers, C. (1981). Conditions for linear processes to be strong-mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(4), 477–480.

Xu, K.-L., & Phillips, P. C. (2008). Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics*, 142(1), 265–280.

Zhang, T., & Wu, W. B. (2012). Inference of time-varying regression models. *Annals of Statistics*, 40(3), 1376–1402.

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