Journal of Econometrics 240 (2024) 105671

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Time-varying multivariate causal processes

Jiti Gao [a], Bin Peng [a,*], Wei Biao Wu [b], Yayi Yan [c]

[a] Department of Econometrics and Business Statistics, Monash University, Australia
[b] Department of Statistics, The University of Chicago, United States of America
[c] School of Statistics and Management, Shanghai University of Finance and Economics, China

ARTICLE INFO

JEL classification:
C14
C32
G15

Keywords:
Local linear quasi-maximum likelihood estimation
Multivariate causal process
Uniform confidence band

ABSTRACT

In this paper, we consider a wide class of time-varying multivariate causal processes that nests many classical and new examples as special cases. We first show the existence of a weakly dependent stationary approximation to initiate our theoretical investigation. We then consider a quasi-maximum likelihood estimation (QMLE), and provide both point-wise and uniform inferences to coefficient functions of interest. The theoretical findings are further examined through extensive simulations. Finally, we show empirical relevance of our study by evaluating both temporal and contemporaneous connectedness between the stock markets of China and U.S.

1. Introduction

Multivariate time series models have been extensively studied and applied over the past a few decades. For example, the family of vector autoregressive (VAR) models is widely used for forecasting macro variables (e.g., Schorfheide and Song, 2015; Berger et al., 2023), modeling policy transmission mechanism (e.g., Miranda-Agrippino and Ricco, 2021; Inoue and Kilian, 2022), measuring connectedness between financial agents (e.g., Diebold and Yilmaz, 2014; Geraci and Gnabo, 2018), etc. Of equivalent importance, the family of multivariate (G)ARCH models has been favored for understanding and predicting temporal dependence of the conditional second moments to facilitate better decision making, so much so that it is heavily used in the literature of asset pricing (e.g., Bollerslev et al., 2020), hedging (e.g., Augustyniak et al., 2023), risk management (e.g., Engle and Siriwardane, 2018), and so forth. To put it in a nutshell, the VAR family usually captures the dynamics by imposing structures on the time series itself, while the (G)ARCH family imposes restrictions on the conditional second moments. Therefore, both the first and second moments of multivariate time series are important from modeling and practical perspectives.

It should be noted that most of the aforementioned studies rely on the stationarity assumption. As Dahlhaus (1997) puts it, the assumption of stationarity guarantees that the increase of the sample size leads to more and more information of the same kind which is basic for an asymptotic theory to make sense. However, stationarity may not always be fulfilled in practice (Preuss et al., 2015; Chen et al., 2022). For example, economic and financial time series data often include different macro shocks and, as a consequence, the behavior may be quite volatile; the climate data may contain certain time trend which recently has attracted considerable attention due to greenhouse emission; etc. In the literature, one maintained assumption is that the coefficients of interest are constant over time, that is, parameter instability is not allowed. Failure to take into account parameter changes, given their presence, always leads to incorrect policy implications and predictions (Bai, 1997).

* Corresponding author.
E-mail address: Bin.Peng@monash.edu (B. Peng).

https://doi.org/10.1016/j.jeconom.2024.105671
Received 6 July 2022; Received in revised form 12 September 2023; Accepted 24 December 2023
Available online 22 January 2024

That said, in order to better capture dynamics of multivariate time series as well as allow for certain nonstationarity,[1] it is reasonable to marry the VAR family and the (G)ARCH family while accounting for parameter instability. In practice, a time-varying VAR setting can be used to study time-variations in policy transmission mechanism or construct a continuously evolving network of spillover effects between financial institutions (Geraci and Gnabo, 2018), while a time-varying GARCH setting can help answer research questions such as: (i) Whether the correlations between asset returns change over time? (ii). Do they increase in the long run, perhaps because of the globalization of financial markets? These questions are of importance for both investors and policymakers (Bauwens et al., 2006; Diebold and Yilmaz, 2009). Despite the large number of potential applications, the statistical tools with the associated asymptotic properties for multivariate time-varying dynamic models are still underdeveloped. In this paper, we aim to contribute along this line, and start by presenting our model below.

Specifically, we consider a class of multivariate causal processes as follows:

$$\mathbf{x}_t = \begin{cases} \boldsymbol{\mu}\left(\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots; \theta(\tau_t)\right) + \mathbf{H}\left(\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots; \theta(\tau_t)\right) \boldsymbol{\varepsilon}_t, & \text{for} \quad t = 1, \ldots, T, \\ \boldsymbol{\mu}\left(\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots; \theta(0)\right) + \mathbf{H}\left(\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots; \theta(0)\right) \boldsymbol{\varepsilon}_t, & \text{for} \quad t \le 0, \end{cases} \tag{1.1}$$

where $\tau_t = t/T$, $\boldsymbol{\mu}\left(\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots; \theta(\tau_t)\right)$ is an m-dimensional vector, $\mathbf{H}\left(\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots; \theta(\tau_t)\right)$ is an $m \times m$ matrix, $\theta(\tau)$ is $d \times 1$ with each element belonging to $C^3[0,1]$, and $\{\boldsymbol{\varepsilon}_t\}$ is a sequence of independent and identically distributed (i.i.d.) random vectors. Both m and d are assumed to be fixed throughout the paper. Notably, both $\boldsymbol{\mu}(\cdot)$ and $\mathbf{H}(\cdot)$ are measurable and have known functional forms. The setup for $t \le 0$ essentially requires $\mathbf{x}_0$ to be stationary, which is commonly adopted in the literature of locally stationary time series analysis (e.g., Section 3.1 of Vogt, 2012). In Section 2.4, we show that models, such as time-varying VARMA/multivariate GARCH/VARMA-GARCH, all admit a multivariate causal representation given by (1.1).

Up to this point, it is worth briefly reviewing the relevant literature. In the regression context, time-varying models of the form, $y_t = \mathbf{z}_t^\top \beta(\tau_t) + e_t$, have been extensively discussed over the past two decade, e.g., Chen and Hong (2012), Zhang and Wu (2012) and Phillips et al. (2017), just to name a few. Since $\mathbf{z}_t$ does not involve any lagged value of y_t, $\mathbf{z}_t$ can be generated by its own structure with stationarity, local stationarity or unit-root depending on the research question. However, for time-varying dynamic models such as (1.1), the process is generated by iterated time-varying random functions, which complicates the development of asymptotic theory. Also, it is difficult to justify that a time-varying dynamic model satisfies widely used (locally) stationary mixing conditions. Indeed, even some stationary AR(1) processes are not necessarily mixing (Doukhan, 2012, Section 2.3.1), unless conditions are imposed on the densities of error terms (Withers, 1981). A common treatment from the literature of locally stationary processes is the so-called "stationary approximation", which usually starts by finding a weakly dependent stationary approximation (e.g., Dahlhaus, 1996 and Zhang and Wu, 2012 on time-varying AR models; Dahlhaus and Polonik, 2009 on time-varying ARMA models; Dahlhaus and Rao, 2006 and Truquet, 2017 on time-varying ARCH models; Karmakar et al., 2022 on time-varying AR-ARCH models). However, this approach relies on the autoregressive representation, and cannot be naturally extended to processes having a more general recursion with infinity memory. This line of research basically focuses on univariate time series.

In view of the aforementioned literature, our first contribution is proposing a wide class of time-varying multivariate causal processes that nests many classical and new examples as special cases. We then show the existence of a weakly dependent stationary approximation for a general class of time-varying multivariate causal processes with infinity memory at any given time of interest (i.e., $\forall \tau \in [0,1]$), which facilitates asymptotic analyses. Second, we establish uniform inference, which can be used to test specific parametric forms. More discussions on the usefulness of uniform inference can also be found in Chen and Christensen (2018), Li and Liao (2020) and Li et al. (2023). In an empirical illustration, we evaluate the time-varying temporal and contemporaneous connectedness between the stock markets of China and U.S. using a time-varying VAR-GARCH model. We find that (1) the contemporaneous interdependence between the two stock markets is statistically significant and is increasing over time, and (2) the temporal return spillovers from U.S. stock market to Chinese stock market significantly vary over time and can be only detected at some local times.

The rest of the paper is organized as follows. Section 2 presents the theoretical findings, while Section 3 provides extensive simulation studies to examine these results. Section 4 investigates the time-varying return/volatility spillovers, as well as time-varying conditional correlations between the Chinese and U.S. stock market. Section 5 concludes. The proofs together with some extra results are given in the online appendices of the paper.

Before proceeding further, it is convenient to introduce some notation: the symbol $|\cdot|$ denotes the Euclidean norm of a vector or the spectral norm for a matrix; $\|\mathbf{v}\|_q := \left(E|\mathbf{v}|^q\right)^{1/q}$ for any $q > 0$ and $\|\cdot\| := \|\cdot\|_2$ for short; $\otimes$ denotes the Kronecker product; $\odot$ denotes the Hadamard product; $\mathbf{I}_a$ stands for an $a \times a$ identity matrix; $\mathbf{0}_{a\times b}$ stands for an $a \times b$ matrix of zeros, and we write $\mathbf{0}_a$ for short when $a = b$; for a function $g(w)$, let $g^{(j)}(w)$ denote the jth derivative of $g(w)$, where $j \ge 0$ and $g^{(0)}(w) \equiv g(w)$; $K_h(\cdot) = K(\cdot/h)/h$, where $K(\cdot)$ and h stand for a nonparametric kernel function and a bandwidth respectively; let $\tilde{c}_k = \int_{-1}^{1} u^k K(u) \mathrm{d}u$ and $\tilde{v}_k = \int_{-1}^{1} u^k K^2(u) \mathrm{d}u$ for integer $k \ge 0$; $\mathrm{diag}(\mathbf{a})$ is a diagonal matrix with the vector $\mathbf{a}$ on its main diagonal, while $\mathrm{diag}(\mathbf{A})$ creates a vector from the diagonal of matrix $\mathbf{A}$; finally, let $\to_P$ and $\to_D$ denote convergence in probability and convergence in distribution, respectively.

[1] Here, we specifically refer to locally stationary type of nonstationarity as in Dahlhaus (1997), in which the author explains why even the simplest AR(1) model with a time-varying parameter is not stationary in length.

2. Estimation and asymptotics

In this section, we first show the existence of a weakly dependent stationary approximation for the model (1.1) in Section 2.1; we then provide the estimation approach using the local linear quasi-maximum-likelihood estimation method and establish the asymptotic properties in Section 2.2; Section 2.3 provides asymptotic properties for both point-wise and uniform inferences; and Section 2.4 gives some detailed examples to showcase our study.

2.1. Stationary approximation

To study (1.1), the first challenge lies in the fact that the process, which is expressed in terms of iterated random functions, may not be stationary. Therefore, we initiate our analysis by finding a stationary approximation of $\{\mathbf{x}_t\}$ for $\forall \tau \in [0,1]$. We are then able to measure the weak dependence of $\{\mathbf{x}_t\}$ using the nonlinear system of Wu (2005), which further enables us to derive the asymptotic properties.

Before proceeding further, we briefly introduce the nonlinear system of Wu (2005). Consider an example in which $\mathbf{e}_t$ is a stationary process, and admits a causal representation $\mathbf{e}_t = \mathbf{J}(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \dots)$ with $\mathbf{J}(\cdot)$ being a measurable function. See Tong (1990, p. 204) for nonlinear time series of this kind. For $k \geq 0$, we define the following dependence measure:

$$\delta_r^{\mathbf{e}}(k) = \left\| \mathbf{J}(\boldsymbol{\varepsilon}_k, \boldsymbol{\varepsilon}_{k-1}, \dots \boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_0, \boldsymbol{\varepsilon}_{-1}, \dots) - \mathbf{J}(\boldsymbol{\varepsilon}_k, \dots, \boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_0^*, \boldsymbol{\varepsilon}_{-1}, \dots) \right\|_r, \tag{2.1}$$

where $\boldsymbol{\varepsilon}_0^*$ is an independent copy of $\{\boldsymbol{\varepsilon}_j\}$, and $\delta_r^{\mathbf{e}}(k)$ quantifies the dependence of $\mathbf{e}_t$ on $\boldsymbol{\varepsilon}_0$.

We now introduce some basic assumptions.

Assumption 1.

1. $\{\boldsymbol{\varepsilon}_t\}$ is a sequence of i.i.d. random vectors with $E(\boldsymbol{\varepsilon}_1) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_1^\top) = \mathbf{I}_m$, and $\|\boldsymbol{\varepsilon}_1\|_r < \infty$ for some $r \geq 2$.
2. For $\forall \mathbf{z}, \mathbf{z}' \in (\mathbb{R}^m)^\infty$ and $\forall \boldsymbol{\vartheta} \in \mathbb{R}^d$, there exist nonnegative sequences $\{\alpha_j(\boldsymbol{\vartheta})\}_{j=1}^\infty$ and $\{\beta_j(\boldsymbol{\vartheta})\}_{j=1}^\infty$ such that
$$|\boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta}) - \boldsymbol{\mu}(\mathbf{z}'; \boldsymbol{\vartheta})| \leq \sum_{j=1}^\infty \alpha_j(\boldsymbol{\vartheta}) |\mathbf{z}_j - \mathbf{z}_j'|,$$
$$|\mathbf{H}(\mathbf{z}; \boldsymbol{\vartheta}) - \mathbf{H}(\mathbf{z}'; \boldsymbol{\vartheta})| \leq \sum_{j=1}^\infty \beta_j(\boldsymbol{\vartheta}) |\mathbf{z}_j - \mathbf{z}_j'|,$$
where $\mathbf{z}_j$ and $\mathbf{z}_j'$ are the jth columns of $\mathbf{z}$ and $\mathbf{z}'$, respectively.
3. For $\forall \tau \in [0,1]$, $\boldsymbol{\theta}(\tau)$ lies in the interior of $\boldsymbol{\Theta}_r$, where
$$\boldsymbol{\Theta}_r := \left\{ \boldsymbol{\vartheta} \in \boldsymbol{\Theta} \mid \sum_{j=1}^\infty \alpha_j(\boldsymbol{\vartheta}) + \|\boldsymbol{\varepsilon}_1\|_r \sum_{j=1}^\infty \beta_j(\boldsymbol{\vartheta}) < 1 \right\},$$
$\boldsymbol{\Theta}$ is a compact set of $\mathbb{R}^d$, and r is the same as that of Assumption 1.1.

Assumption 1.1 is standard when studying dynamic time series model (Lütkepohl, 2005, p. 563). Assumption 1.2 imposes Lipschitz-type conditions on $\boldsymbol{\mu}(\cdot)$ and $\mathbf{H}(\cdot)$, which are rather minor, and can be easily fulfilled by a variety of models such as those to be studied in Section 2.4. Assumption 1.3 guarantees a weakly dependent stationary approximation of $\mathbf{x}_t$ for $\forall \tau \in [0,1]$, and can be considered as a nonparametric extension to that adopted in Bardet and Wintenberger (2009). Assumption 1 ensures that we can study many multivariate dynamic models in a unified framework.

We now present the following proposition which facilitates the development.

Proposition 2.1. *Let Assumption 1 hold. For any $\tau \in [0,1]$, there exists a stationary process*

$$\widetilde{\mathbf{x}}_t(\tau) = \boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{t-1}(\tau), \widetilde{\mathbf{x}}_{t-2}(\tau), \dots; \boldsymbol{\theta}(\tau)\right) + \mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau), \widetilde{\mathbf{x}}_{t-2}(\tau), \dots; \boldsymbol{\theta}(\tau)\right) \boldsymbol{\varepsilon}_t$$

such that

1. *$\widetilde{\mathbf{x}}_t(\tau)$ admits a causal representation $\widetilde{\mathbf{x}}_t(\tau) = \mathbf{J}(\tau, \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \dots)$ with $\mathbf{J}(\cdot)$ being a measurable function, and $\sup_{\tau \in [0,1]} \|\widetilde{\mathbf{x}}_t(\tau)\|_r < \infty$,*
2. *$\delta_r^{\widetilde{\mathbf{x}}(\tau)}(k) \leq O(1) \inf_{1 \leq p \leq k} \{\rho(\tau)^{k/p} + \sum_{j=p+1}^\infty [\alpha_j(\boldsymbol{\theta}(\tau)) + \beta_j(\boldsymbol{\theta}(\tau))]\} \to 0$ as $k \to \infty$,*

where $\|\cdot\|_r$ is defined in the end of Section 1, and $\rho(\tau) := \sum_{j=1}^\infty \alpha_j(\boldsymbol{\theta}(\tau)) + \|\boldsymbol{\varepsilon}_1\|_r \sum_{j=1}^\infty \beta_j(\boldsymbol{\theta}(\tau))$.

For a univariate p-Markov process of the form

$$\widetilde{x}_{p,t}(\tau) = \sum_{i=1}^p \alpha_i(\tau) m_i(\widetilde{x}_{t-i}(\tau)) + \left(\sum_{i=0}^p \beta_i(\tau) v_i(\widetilde{x}_{t-i}(\tau)) \right)^{1/2} \varepsilon_t$$

with some known functions $\{m_i(\cdot)\}$ and $\{v_i(\cdot)\}$, Karmakar et al. (2022) show that there exists $0 < \rho < 1$ such that $\sup_{\tau \in [0,1]} \delta_r^{\widetilde{x}_p(\tau)}(k) = O(\rho^k)$ based on the development of Wu and Shao (2004). Their paper rules out the time-varying versions of some widely used models (e.g., ARMA, GARCH, ARMA-GARCH).

From a methodological viewpoint, we give a set of new proofs which allows us to measure the dependence of multivariate causal processes with infinite memory. Here, the concept of infinite memory refers to the idea that the future state of the process depends on an infinite history of its past states. Thus, the dependence $\delta_r^{\widetilde{\mathbf{x}}(\tau)}(k)$ relies on the choice of p and the decay rates of the coefficients $\alpha_j(\theta(\tau))$ and $\beta_j(\theta(\tau))$.

To ensure that $\widetilde{\mathbf{x}}_t(\tau)$ can approximate $\mathbf{x}_t$ reasonably well, we impose more structure.

Assumption 2.

1. There exists a nonnegative sequence $\{\chi_j\}$ with $\sum_{j=1}^{\infty}\chi_j < \infty$ such that for $\forall \mathbf{z} \in (\mathbb{R}^m)^\infty$ and $\forall \boldsymbol{\vartheta}, \boldsymbol{\vartheta}' \in \boldsymbol{\Theta}_r$

$$|\boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta}) - \boldsymbol{\mu}(\mathbf{z}; \boldsymbol{\vartheta}')| + |\mathbf{H}(\mathbf{z}; \boldsymbol{\vartheta}) - \mathbf{H}(\mathbf{z}; \boldsymbol{\vartheta}')| \le |\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'| \sum_{j=1}^{\infty} \chi_j |\mathbf{z}_j|.$$

2. Let $\sup_{\tau\in[0,1]} \alpha_j(\theta(\tau)) = O(j^{-(2+s)})$ and $\sup_{\tau\in[0,1]} \beta_j(\theta(\tau)) = O(j^{-(2+s)})$ for some $s > 0$ as $j \to \infty$.

Assumption 2.1 imposes another Lipschitz-type condition with respect to the parameter space. Assumption 2.2 further regulates the decay rates of $\alpha_j(\theta(\tau))$ and $\beta_j(\theta(\tau))$.

Using Assumptions 1 and 2, we may quantify the distance between $\widetilde{\mathbf{x}}_t(\tau)$ and $\mathbf{x}_t$ as follows.

Proposition 2.2. *Suppose Assumptions 1 and 2 hold. Then*

1. $\|\widetilde{\mathbf{x}}_1(\tau) - \widetilde{\mathbf{x}}_1(\tau')\|_r = O(|\tau - \tau'|)$ *for* $\forall \tau, \tau' \in [0, 1]$,
2. $\max_{t\ge 1} \|\mathbf{x}_t - \widetilde{\mathbf{x}}_t(\tau_t)\|_r = O(T^{-1})$,

where $\|\cdot\|_r$ *is defined in the end of Section 1.*

Proposition 2.2 can be considered as the stochastic version of the Hölder continuity. We are now ready to investigate the estimation theory in the next subsection.

2.2. Estimation

Since $\boldsymbol{\varepsilon}_t$ may not be normally distributed, we consider the local linear quasi-maximum-likelihood estimation (QMLE) method (Fan and Gijbels, 1996, p. 194). The main idea is to pretend $\boldsymbol{\varepsilon}_t$ is normally distributed, so we can have the conditional likelihood as follows:

$$\frac{1}{T}\sum_{t=1}^{T} \ell(\mathbf{x}_t, \mathbf{z}_{t-1}; \boldsymbol{\theta}(\tau_t)),$$

where $\mathbf{z}_t = (\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots)$, and

$$\begin{aligned} &\ell(\mathbf{x}_t, \mathbf{z}_{t-1}; \boldsymbol{\vartheta}) \\ =& -\frac{1}{2}(\mathbf{x}_t - \boldsymbol{\mu}(\mathbf{z}_{t-1}; \boldsymbol{\vartheta}))^\top \left(\mathbf{H}(\mathbf{z}_{t-1}; \boldsymbol{\theta}(\tau_t))\mathbf{H}(\mathbf{z}_{t-1}; \boldsymbol{\vartheta})^\top\right)^{-1} (\mathbf{x}_t - \boldsymbol{\mu}(\mathbf{z}_{t-1}; \boldsymbol{\vartheta})) \\ & -\frac{1}{2}\log\det\left(\mathbf{H}(\mathbf{z}_{t-1}; \boldsymbol{\theta}(\tau_t))\mathbf{H}(\mathbf{z}_{t-1}; \boldsymbol{\vartheta})^\top\right). \end{aligned}$$

From a nonparametric modeling perspective, it suffices to consider the data points in the neighborhood around the given time point τ (i.e., $|\tau_t - \tau| \le h$) when estimating $\boldsymbol{\theta}(\tau)$, as we have

$$\boldsymbol{\theta}(\tau_t) \simeq \boldsymbol{\theta}(\tau) + h\boldsymbol{\theta}^{(1)}(\tau)\cdot\frac{\tau_t - \tau}{h}.$$

Therefore, we are able to parameterize $\boldsymbol{\theta}(\tau_t)$. Also, in practice, we only observe $\mathbf{x}_t$ for $t \ge 1$, so we have to work with the truncated version of $\mathbf{z}_t$ for each $t \ge 1$:

$$\mathbf{z}_t^c = (\mathbf{x}_t, \ldots, \mathbf{x}_1, \mathbf{0}, \ldots).$$

Finally, our likelihood function is specified as follows:

$$\mathcal{L}_\tau(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2) = \frac{1}{T}\sum_{t=1}^{T} \ell(\mathbf{x}_t, \mathbf{z}_{t-1}^c; \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_t - \tau)/h) K_h(\tau_t - \tau). \tag{2.2}$$

Accordingly, for $\forall \tau$, $(\boldsymbol{\theta}(\tau), h\boldsymbol{\theta}^{(1)}(\tau))$ are estimated by

$$(\widehat{\boldsymbol{\theta}}(\tau), \widehat{\boldsymbol{\theta}}^\star(\tau)) = \underset{(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)\in \mathbf{E}_T(r)}{\operatorname{argmax}}\ \mathcal{L}_\tau(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2), \tag{2.3}$$

where $\mathbf{E}_T(r) = \boldsymbol{\Theta}_r \times (h \cdot \boldsymbol{\Theta}^{(1)})$, $\boldsymbol{\Theta}^{(1)}$ is a compact set and $\widehat{\boldsymbol{\theta}}^\star(\tau)$ denotes the estimator of $h\boldsymbol{\theta}^{(1)}(\tau)$.

We impose more assumptions in order to derive the asymptotic distribution.

Assumption 3.

1. $\inf_{\vartheta\in\Theta_r, \mathbf{z}\in(\mathbb{R}^m)^\infty} \lambda_{\min}\left(\mathbf{H}(\mathbf{z};\vartheta)\mathbf{H}(\mathbf{z};\vartheta)^\top\right) \geq \underline{c}$ for some $\underline{c} > 0$, where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of a square matrix.
2. For any $\vartheta \in \Theta_r$, $\mu(\widetilde{\mathbf{z}}_t(\tau);\theta(\tau)) = \mu(\widetilde{\mathbf{z}}_t(\tau);\vartheta)$ and $\mathbf{H}(\widetilde{\mathbf{z}}_t(\tau);\theta(\tau)) = \mathbf{H}(\widetilde{\mathbf{z}}_t(\tau);\vartheta)$ a.s. imply $\vartheta = \theta(\tau)$ for some τ, where $\widetilde{\mathbf{z}}_t(\tau) = \left(\widetilde{\mathbf{x}}_t(\tau), \widetilde{\mathbf{x}}_{t-1}(\tau), \ldots\right)$.

Assumption 4.

1. $\mu(\cdot;\vartheta)$ and $\mathbf{H}(\cdot;\vartheta)$ are twice continuously differentiable with respect to ϑ.
2. There exists a nonnegative sequence $\{\chi_j\}_{j=1}^\infty$ with $\chi_j = O(j^{-(2+s)})$ and some $s > 0$ such that for any $\mathbf{z},\mathbf{z}' \in (\mathbb{R}^m)^\infty$ and any $\vartheta,\vartheta' \in \Theta_r$:

$$|\nabla_\vartheta^k \mu(\mathbf{z};\vartheta) - \nabla_\vartheta^k \mu(\mathbf{z};\vartheta')| + |\nabla_\vartheta^k \mathbf{H}(\mathbf{z};\vartheta) - \nabla_\vartheta^k \mathbf{H}(\mathbf{z};\vartheta')| \leq |\vartheta - \vartheta'| \sum_{j=1}^\infty \chi_j |\mathbf{z}_j|,$$

$$|\nabla_\vartheta^k \mu(\mathbf{z};\vartheta) - \nabla_\vartheta^k \mu(\mathbf{z}';\vartheta)| + |\nabla_\vartheta^k \mathbf{H}(\mathbf{z};\vartheta) - \nabla_\vartheta^k \mathbf{H}(\mathbf{z}';\vartheta)| \leq \sum_{j=1}^\infty \chi_j |\mathbf{z}_j - \mathbf{z}_j'|,$$

where $\nabla_\vartheta = \left(\frac{\partial}{\partial\vartheta_1}, \ldots, \frac{\partial}{\partial\vartheta_d}\right)^\top$, and $k = 1, 2$.

Assumption 5. Let $K(\cdot)$ be a symmetric and positive kernel function defined on $[-1,1]$ with $\int_{-1}^1 K(u)\mathrm{d}u = 1$. Moreover, $K(\cdot)$ is Lipschitz continuous on $[-1,1]$. As $(T,h) \to (\infty,0)$, $Th \to \infty$.

Assumption 3.1 ensures the positive definiteness of the covariance matrix of the likelihood function, and is widely adopted when studying multivariate time series (e.g., Bardet and Wintenberger, 2009, p. 2736). In fact, the validity of this assumption is easy to justify in view of (2.8), (2.12) and (2.15) for Examples 1–3 to be studied in Section 2.4 below. Assumption 3.2 imposes a standard identification condition in the literature of M-estimation (e.g., Proposition 3.4 of Jeantheau, 1998). It is noteworthy that the current form of Assumption 3 accommodates the flexibility of Model (1.1), which in fact may be unnecessary if we have a detailed model in practice. See Section 2.4 for example.

Assumption 4 imposes Lipschitz-type conditions on the first and second order derivatives of $\mu(\cdot)$ and $\mathbf{H}(\cdot)$ to ensure the smoothness of their functional components.

Assumption 5 is a set of mild conditions on the kernel function and the bandwidth.

With these conditions in hand, we summarize the first theorem below.

Theorem 2.1. *Suppose Assumption 1 with $r \geq 6$ and Assumptions 2–5 hold.*

(1). If $Th^7 \to 0$, then for any $\tau \in (0,1)$

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

where $\Sigma_\theta(\tau) = \Sigma^{-1}(\tau)\Omega(\tau)\Sigma^{-1}(\tau)$, $\Sigma(\tau) = E\left(\nabla_\vartheta^2 \ell(\widetilde{\mathbf{x}}_1(\tau), \widetilde{\mathbf{z}}_0(\tau);\theta(\tau))\right)$ and

$$\Omega(\tau) = E\left(\nabla_\vartheta \ell(\widetilde{\mathbf{x}}_1(\tau), \widetilde{\mathbf{z}}_0(\tau);\theta(\tau)) \cdot \nabla_\vartheta \ell(\widetilde{\mathbf{x}}_1(\tau), \widetilde{\mathbf{z}}_0(\tau);\theta(\tau))^\top\right).$$

(2). In addition, if ε_t is normally distributed, we have $\Omega(\tau) = -\Sigma(\tau)$ and thus $\Sigma_\theta(\tau) = \Omega^{-1}(\tau)$.

Note that the condition $Th^7 \to 0$ is less restrictive than $Th^5 \to 0$, so $Th^7 \to 0$ does not indicate the bias term will vanish in general. If $Th^5 \to 0$, there is no bias term involved in the asymptotic distribution of Theorem 2.1, which then falls in the usual undersmoothing scenario (Li and Racine, 2007, p. 15). In general, in order to establish valid inferences, both $\theta^{(2)}(\tau)$ and $\Sigma_\theta(\tau)$ have to be accounted for (cf., Xia, 1998). In the following subsection, we shall establish both point-wise inference and uniform inference after dealing with the bias term.

Based on Theorem 2.1, we can give the rate of forecast error of k-step ahead predictions for (1.1), which is one of the most important tasks of time series analysis (cf., Stock and Watson, 2001). For any fixed integer $k \geq 1$, define the forecasts of the conditional mean and conditional variance of (1.1) by

$$\widehat{\mathbf{x}}_{T+k|T} = \mu\left(\widehat{\mathbf{x}}_{T+k-1|T}, \widehat{\mathbf{x}}_{T+k-2|T}, \ldots; \widehat{\theta}(1)\right)$$

and

$$\widehat{\mathbf{M}}_{T+k|T} = \mathbf{M}\left(\widehat{\mathbf{x}}_{T+k-1|T}, \widehat{\mathbf{x}}_{T+k-2|T}, \ldots; \widehat{\theta}(1)\right),$$

where $\widehat{\mathbf{x}}_{t|T} = \mathbf{x}_t$ for $1 \leq t \leq T$, $\widehat{\mathbf{x}}_{t|T} = \mathbf{0}$ for $t \leq 0$ and $\mathbf{M}(\mathbf{z};\vartheta) = \mathbf{H}(\mathbf{z};\vartheta)\mathbf{H}^\top(\mathbf{z};\vartheta)$. We then have the following corollary.

Corollary 2.1. *Under the conditions of Theorem 2.1.1, and assume $\theta(\tau)$ to be Lipschitz continuous for $\tau \geq 1$, then for any fixed integer $k \geq 1$*

$$\widehat{\mathbf{x}}_{T+k|T} - \mathbf{x}_{T+k|T} = O_P\left(h^2 + 1/\sqrt{Th}\right)$$

and

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

where $\mathbf{x}_{T+k|T}$ *is recursively defined as* $\mathbf{x}_{T+k|T} = \boldsymbol{\mu}\left(\mathbf{x}_{T+k-1|T}, \mathbf{x}_{T+k-2|T}, \dots; \theta(\tau_{T+k})\right)$ *for* $k \geq 1$, $\mathbf{x}_{t|T} = \mathbf{x}_t$ *for* $t \leq T$, *and*

$$\mathbf{M}_{T+k|T} = \mathbf{H}\left(\mathbf{x}_{T+k-1|T}, \mathbf{x}_{T+k-2|T}, \dots; \theta(\tau_{T+k})\right) \mathbf{H}^\top\left(\mathbf{x}_{T+k-1|T}, \mathbf{x}_{T+k-2|T}, \dots; \theta(\tau_{T+k})\right).$$

2.3. Inference

In this section, we first discuss how to conduct point-wise inference, and then move on to derive the asymptotic results associated with the uniform inference. Specifically, for $\forall \alpha \in (0,1)$, we shall construct a $100(1-\alpha)\%$ asymptotic uniform confidence band (UCB) $\{\Upsilon(\tau), 0 \leq \tau \leq 1\}$ for $\theta(\cdot)$ in the sense that

$$\lim_{T\to\infty} \Pr\left(\theta(\tau) \in \Upsilon(\tau), 0 \leq \tau \leq 1\right) = 1 - \alpha.$$

Notably, the uniform inference nests the traditional constancy test as a special case. It does not only allow one to examine whether a time-varying model should be preferred to its parametric counterpart, but also allows one to test any particular functional form of interest. For example, if a horizontal line can be embedded in the UCB $\{\Upsilon(\tau)\}$, then we accept the hypothesis that some elements of $\theta(\tau)$ are constant.

Point-wise Inference — First, we construct a jackknife bias-corrected estimator in order to remove the asymptotic bias of Theorem 2.1. Specifically, we let

$$\widetilde{\theta}(\tau) = 2\widehat{\theta}_{h/\sqrt{2}}(\tau) - \widehat{\theta}(\tau), \tag{2.4}$$

where $\widehat{\theta}_{h/\sqrt{2}}(\tau)$ is defined in the same way as $\widehat{\theta}(\tau)$ but using the bandwidth $h/\sqrt{2}$.

After some tedious development (Lemma B.7 of Appendix B), we have uniformly over $\tau \in [h, 1-h]$

$$\begin{aligned}\widetilde{\theta}(\tau) - \theta(\tau) \;=\; & -\boldsymbol{\Sigma}^{-1}(\tau)\frac{1}{Th}\sum_{t=1}^{T}\widetilde{K}((\tau_t-\tau)/h)\nabla_{\vartheta}\ell(\widetilde{\mathbf{x}}_t(\tau_t), \widetilde{\mathbf{z}}_{t-1}(\tau_t); \theta(\tau_t)) \\ & + O_P((Th)^{-1/2}h^{3/2}(\log T)^{1/2}) + o(h^3),\end{aligned}$$

where $\widetilde{K}(x) = 2\sqrt{2}K(\sqrt{2}x) - K(x)$ that is essentially a fourth-order kernel. It then infers that under the conditions of Theorem 2.1,

$$\sqrt{Th}(\widetilde{\theta}(\tau) - \theta(\tau)) \to_D N\left(\mathbf{0}, v_0\boldsymbol{\Sigma}_{\theta}(\tau)\right),$$

where $v_0 = \int_{-1}^{1}\widetilde{K}^2(u)\mathrm{d}u$.

It is noteworthy that the construction of (2.4) is different from directly using the fourth-order kernel in the regression. In terms of bandwidth selection, the traditional methods (e.g., cross-validation) still remain valid for (2.4) (Richter and Dahlhaus, 2019). However, if one directly employs the fourth-order kernel in the regression, it remains unclear how to select the optimal bandwidth in practice.

Now we discuss how to estimate $\boldsymbol{\Sigma}_{\theta}(\tau)$ which is constructed by $\boldsymbol{\Sigma}(\tau)$ and $\boldsymbol{\Omega}(\tau)$. Intuitively, we consider the following estimator

$$\widehat{\boldsymbol{\Sigma}}_{\theta}(\tau) = \widehat{\boldsymbol{\Sigma}}^{-1}(\tau)\widehat{\boldsymbol{\Omega}}(\tau)\widehat{\boldsymbol{\Sigma}}^{-1}(\tau), \tag{2.5}$$

where

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}(\tau) &= A_T(\tau)^{-1}\sum_{t=1}^{T}\nabla_{\vartheta}^2\ell(\mathbf{x}_t, \mathbf{z}_{t-1}^c; \widehat{\theta}(\tau))K_h(\tau_t-\tau), \\ \widehat{\boldsymbol{\Omega}}(\tau) &= A_T(\tau)^{-1}\sum_{t=1}^{T}\nabla_{\vartheta}\ell(\mathbf{x}_t, \mathbf{z}_{t-1}^c; \widehat{\theta}(\tau)) \cdot \nabla_{\vartheta}\ell(\mathbf{x}_t, \mathbf{z}_{t-1}^c; \widehat{\theta}(\tau))^\top K_h(\tau_t-\tau), \\ A_T(\tau) &= \sum_{t=1}^{T}K_h(\tau_t-\tau).\end{aligned}$$

The following corollary summarizes the asymptotic property of (2.5).

Corollary 2.2. *Under the conditions of Theorem 2.1.1, suppose further that*

$$\sup_{\tau\in[0,1]}[\alpha_j(\theta(\tau)) + \beta_j(\theta(\tau))] = O(j^{-(5/2+s)})$$

for some $s > 0$. *In addition, let* $h(\log T)^2 \to 0$ *and* $T^{1-6/r}h \to \infty$. *Then*

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

Uniform Inference — We now consider the uniform inference. To allow for flexibility, we first introduce a selection matrix $\mathbf{C}$ with full row rank, which selects the parameters of interest as follows:

$$\theta_{\mathbf{C}}(\tau) := \mathbf{C}\theta(\tau).$$

Accordingly, the estimator and the corresponding asymptotic covariance matrix become

$$\widehat{\theta}_{\mathbf{C}}(\tau) := \mathbf{C}\widehat{\theta}(\tau) \quad \text{and} \quad \Sigma_{\mathbf{C}}(\tau) = \mathbf{C}\Sigma_{\theta}(\tau)\mathbf{C}^{\top}.$$

Theorem 2.2. *Under the conditions of Theorem 2.1.1, suppose further that*

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

for some $s > 0$. *In addition, let* $(\log T)^4/(T^{\nu}h) \to 0$ *with* $\nu = \frac{1}{2} - \frac{r-6}{4rs/3+2r-4}$ *and* $Th^7\log T \to 0$. *Then for all* $u \in \mathbb{R}$

$$\lim_{T\to\infty}\Pr\left(\sqrt{\frac{Th}{\widetilde{v}_0}}\sup_{\tau\in[h,1-h]}\left|\Sigma_{\mathbf{C}}^{-1/2}(\tau)\left\{\widehat{\theta}_{\mathbf{C}}(\tau) - \theta_{\mathbf{C}}(\tau) - \frac{1}{2}h^2\widetilde{c}_2\theta_{\mathbf{C}}^{(2)}(\tau)\right\}\right| - B(1/h) \le \frac{u}{\sqrt{2\log(1/h)}}\right) = \exp(-2\exp(-u)),$$

where

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

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

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

In Theorem 2.2, ν is slightly smaller than $1/2$ as we only require r to be slightly larger than 6. Hence, the usual optimal bandwidth $h_{opt} = O(T^{-1/5})$ satisfies the conditions $(\log T)^4/(T^{\nu}h) \to 0$ and $Th^7\log T \to 0$. Theorem 2.2 states that the maximum deviation of $\widehat{\theta}(\tau) - \theta(\tau)$ converges to a Gumbel distribution after suitable normalization. To provide more insights on Theorem 2.2, we briefly outline its proof strategy. By Lemma B.7 in Appendix B, we have the Bahadur representation uniformly over $\tau \in [h, 1-h]$:

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

Furthermore, by using Gaussian approximation technique in Lemma B.10, we have

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

where $\{\mathbf{v}_t\}$ is a sequence of i.i.d. standard normal vectors. Then, Theorem 2.2 follows from the fact that the maximum of i.i.d. standard normal variables converges to the standard Gumbel distribution according to the extreme value theory of Gaussian processes.

Note that the approximation rate of Gumbel distribution for the maximum of i.i.d. normal variables is $1/\log T$ (cf., Hall (1980)), and thus the convergence rate of the UCB for $\theta_{\mathbf{C}}(\cdot)$ is extremely slow. Based on the Gaussian approximation results as stated above, it is possible improve the approximation rate of Gumbel distribution when constructing UCB of $\theta(\cdot)$. Here, we consider a bootstrap method to improve the finite sample performance. We summarize the result in the following corollary.

Corollary 2.3. *Under the conditions of Theorem 2.2. Suppose that* $h = O(T^{-\kappa})$ *with* $1/7 < \kappa < \nu$. *Then, on a richer probability space, there exists i.i.d.* k*-dimensional standard normal variables* $\mathbf{v}_1, \dots, \mathbf{v}_T$ *such that*

$$\sup_{\tau\in[0,1]}|\widehat{\theta}_{\mathbf{C}}(\tau) - \theta_{\mathbf{C}}(\tau) - \frac{1}{2}h^2b_h(\tau)\theta_{\mathbf{C}}^{(2)}(\tau) - \Sigma_{\mathbf{C}}^{1/2}(\tau)\mathbf{V}_h^*(\tau)| = O_P\left(\frac{T^{-\alpha}}{\sqrt{Th\log T}}\right),$$

where $\alpha = \min\{(\nu-\kappa)/2, (7\kappa-1)/2, \kappa/2\}$, $\widetilde{c}_{k,h}(\tau) = \int_{-\tau/h}^{(1-\tau)/h}u^kK(u)\mathrm{d}u$, $\mathbf{V}_h^*(\tau) = T^{-1}\sum_{t=1}^{T}\mathbf{v}_t\omega_{t,h}(\tau)$,

$$b_h(\tau) = \frac{\widetilde{c}_{2,h}^2(\tau) - \widetilde{c}_{1,h}(\tau)\widetilde{c}_{3,h}(\tau)}{\widetilde{c}_{0,h}(\tau)\widetilde{c}_{2,h}(\tau) - \widetilde{c}_{1,h}^2(\tau)} \quad \text{and} \quad \omega_{t,h}(\tau) = K_h(\tau_t - \tau)\frac{\widetilde{c}_{2,h}(\tau) - \frac{\tau_t-\tau}{h}\widetilde{c}_{1,h}(\tau)}{\widetilde{c}_{0,h}(\tau)\widetilde{c}_{2,h}(\tau) - \widetilde{c}_{1,h}^2(\tau)}.$$

By Corollary 2.3, we propose the following numerical procedure to construct the UCB of $\theta_{\mathbf{C}}(\tau)$:

Step 1 Use the sample $\{\mathbf{x}_t\}_{t=1}^T$ to estimate $\widehat{\theta}_{\mathbf{C}}(\tau)$ by (2.3), and compute $\widetilde{\theta}_{\mathbf{C}}(\tau)$ based on (2.4).

Step 2 Generate i.i.d. k-dimensional standard normal variables $\{\mathbf{v}_t^*\}$ and calculate the quantity $\sup_{\tau\in[0,1]}|\mathbf{V}_h^*(\tau)|$, in which $\mathbf{V}_h^*(\tau) = T^{-1}\sum_{t=1}^T \mathbf{v}_t^*(2\omega_{t,h/\sqrt{2}}(\tau) - \omega_{t,h}(\tau))$.

Step 3 Repeat Step 2 R times to obtain the empirical $(1-\alpha)$th quantile $\widehat{q}_{1-\alpha}$ of $\sup_{\tau\in[0,1]}|\mathbf{V}_h^*(\tau)|$.

Step 4 Calculate $\widehat{\boldsymbol{\Sigma}}_{\mathbf{C}}(\tau)$ using (2.5), and construct the UCB of $\theta_{\mathbf{C}}(\tau)$ by $\widetilde{\theta}_{\mathbf{C}}(\tau) + \widehat{\boldsymbol{\Sigma}}_{\mathbf{C}}^{1/2}(\tau)\widehat{q}_{1-\alpha}\mathbb{B}_k$, where $\mathbb{B}_k = \{\mathbf{u}\in\mathbb{R}^k : |\mathbf{u}| \le 1\}$ is the unit ball, and k is the rank of $\mathbf{C}$.

Note that for $\tau \in [h, 1-h]$ and $\tau \in \{0,1\}$ we have $b_{h/\sqrt{2}}(\tau) - b_h(\tau) = 0$, and thus the bias term disappears. But for $\tau \in (0,h)\cup(1-h,1)$, we have $b_{h/\sqrt{2}}(\tau) - b_h(\tau) \neq 0$ but $|b_{h/\sqrt{2}}(\tau) - b_h(\tau)| < |b_h(\tau)|$, that is, the jack-knife estimator can only reduce the bias to some extent at the boundary points. Even this is the case, we can still establish UCB as in Corollary 2.3.

2.4. Examples

Below, we demonstrate the usefulness of the aforementioned results by considering Examples 1–3 below. We refer interested readers to Ling (2003), Ling and McAleer (2003) and Bardet and Wintenberger (2009) for extensive investigations on the parametric counterparts of these examples.

Example 1. Consider a time-varying VARMA(p,q) model

$$\mathbf{x}_t = \mathbf{a}(\tau_t) + \sum_{j=1}^p \mathbf{A}_j(\tau_t)\mathbf{x}_{t-j} + \boldsymbol{\eta}_t + \sum_{j=1}^q \mathbf{B}_j(\tau_t)\boldsymbol{\eta}_{t-j}, \quad \boldsymbol{\eta}_t = \boldsymbol{\Omega}^{1/2}(\tau_t)\boldsymbol{\varepsilon}_t, \quad t = 1,2,\dots,T. \tag{2.6}$$

For $\forall\tau \in [0,1]$, simple algebra shows that

$$\widetilde{\mathbf{x}}_t(\tau) = \boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{t-1}(\tau), \widetilde{\mathbf{x}}_{t-2}(\tau), \dots; \theta(\tau)\right) + \mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau), \widetilde{\mathbf{x}}_{t-2}(\tau), \dots; \theta(\tau)\right)\boldsymbol{\varepsilon}_t, \tag{2.7}$$

where

$$\begin{aligned}
&\theta(\tau) = [\mathrm{vec}(\mathbf{a}(\tau), \mathbf{A}_1(\tau), \dots, \mathbf{A}_p(\tau), \mathbf{B}_1(\tau), \dots, \mathbf{B}_q(\tau)); \mathrm{vech}(\boldsymbol{\Omega}(\tau))],\\
&\boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{t-1}(\tau), \widetilde{\mathbf{x}}_{t-2}(\tau), \dots; \theta(\tau)\right) = \mathbf{B}_\tau^{-1}(1)\mathbf{a}(\tau) + \sum_{j=1}^\infty \boldsymbol{\Gamma}_j(\tau)\widetilde{\mathbf{x}}_{t-j}(\tau),\\
&\mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau), \widetilde{\mathbf{x}}_{t-2}(\tau), \dots; \theta(\tau)\right) = \boldsymbol{\Omega}^{1/2}(\tau).
\end{aligned} \tag{2.8}$$

Additionally, in (2.8), $\boldsymbol{\Gamma}_j(\tau)$ is yielded as follows:

$$\mathbf{I}_m - \sum_{j=1}^\infty \boldsymbol{\Gamma}_j(\tau)L^j := \mathbf{B}_\tau^{-1}(L)\mathbf{A}_\tau(L),$$

where $\mathbf{A}_\tau(L) := \mathbf{I}_m - \mathbf{A}_1(\tau)L - \cdots - \mathbf{A}_p(\tau)L^p$ and $\mathbf{B}_\tau(L) := \mathbf{I}_m + \mathbf{B}_1(\tau)L + \cdots + \mathbf{B}_q(\tau)L^q$.

Then we are able to present the following proposition.

Proposition 2.3. *Let $\|\varepsilon_t\|_r < \infty$ for some $r > 4$. Suppose that there is a compact set*

$$\boldsymbol{\Theta} = \{\boldsymbol{\vartheta} = [\mathrm{vec}(\mathbf{a}, \mathbf{A}_1, \dots, \mathbf{A}_p, \mathbf{B}_1, \dots, \mathbf{B}_q); \mathrm{vech}(\boldsymbol{\Omega})] \mid \boldsymbol{\vartheta} \in \mathbb{R}^d\}$$

such that (1). for $\forall\tau \in [0,1]$, $\theta(\tau)$ lies in the interior of $\boldsymbol{\Theta}$, (2). $\det(\mathbf{A}(L)\mathbf{B}(L)) \neq 0$ for all $|L| \le 1$, where $\mathbf{A}(L) := \mathbf{I}_m - \mathbf{A}_1L - \cdots - \mathbf{A}_pL^p$ and $\mathbf{B}(L) := \mathbf{I}_m + \mathbf{B}_1L + \cdots + \mathbf{B}_qL^q$ are coprime and satisfy some necessary identification conditions, (3). $\boldsymbol{\Omega} > 0$. Then, the results of Theorems 2.1 and 2.2 hold for the model (2.6).

We note that the detailed identification conditions required for VARMA processes (e.g., the final equations form or echelon form) can be found in Lütkepohl (2005, p. 452), so we no longer discuss them here in order not to deviate from our main goal.

Due to the specific structure of VARMA processes, we are able to provide the estimation and inferential methods for time-varying structural impulse response, which are of interest in macroeconometrics. Here, we study the impulse response subject to both short-run timing and long-run restrictions. The economic interpretations of the two types of identification conditions can be found in Kilian and Lütkepohl (2017, p. 213 and p. 265). Our results extend their methods by allowing for time-variations in these impulse responses. The results are summarized in Appendix A.2 of the online supplementary document.

Example 2. Consider a time-varying multivariate GARCH(p,q) model

$$\begin{aligned}
\mathbf{x}_t &= \mathrm{diag}(h_{1,t}^{1/2}, \dots, h_{m,t}^{1/2})\boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t = \boldsymbol{\Omega}^{1/2}(\tau_t)\boldsymbol{\varepsilon}_t, \quad t = 1,2,\dots,T,\\
\mathbf{h}_t &= \mathbf{c}_0(\tau_t) + \sum_{j=1}^p \mathbf{C}_j(\tau_t)\left(\mathbf{x}_{t-j}\odot\mathbf{x}_{t-j}\right) + \sum_{j=1}^q \mathbf{D}_j(\tau_t)\mathbf{h}_{t-j},
\end{aligned} \tag{2.9}$$

where $h_{j,t}$ stands for the jth element of $\mathbf{h}_t$, $\odot$ denotes the Hadamard product and

$$\boldsymbol{\Omega}(\tau)=\begin{bmatrix} 1 & \rho_{1,2}(\tau) & \cdots & \rho_{1,m}(\tau) \\ \rho_{1,2}(\tau) & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{m-1,m}(\tau) \\ \rho_{1,m}(\tau) & \rho_{m-1,m}(\tau) & \ddots & 1 \end{bmatrix}. \tag{2.10}$$

For $\forall \tau \in [0,1]$,

$$\widetilde{\mathbf{x}}_t(\tau)=\mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau),\widetilde{\mathbf{x}}_{t-2}(\tau),\dots;\theta(\tau)\right)\boldsymbol{\varepsilon}_t, \tag{2.11}$$

where $\theta(\tau)=[\mathrm{vec}(\mathbf{c}_0(\tau),\mathbf{C}_1(\tau),\dots,\mathbf{C}_p(\tau),\mathbf{D}_1(\tau),\dots,\mathbf{D}_q(\tau));\mathrm{vechl}(\boldsymbol{\Omega}(\tau))]$, the operator $\mathrm{vechl}(\cdot)$ stacks the lower triangular part of a matrix excluding the diagonal and

$$\begin{aligned}&\mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau),\widetilde{\mathbf{x}}_{t-2}(\tau),\dots;\theta(\tau)\right)\\ &=\mathrm{diag}^{1/2}\left(\mathbf{D}_\tau^{-1}(1)\mathbf{c}_0(\tau)+\sum_{j=1}^{\infty}\boldsymbol{\Psi}_j(\tau)\left(\widetilde{\mathbf{x}}_{t-j}(\tau)\odot\widetilde{\mathbf{x}}_{t-j}(\tau)\right)\right)\boldsymbol{\Omega}^{1/2}(\tau).\end{aligned} \tag{2.12}$$

Note that $\boldsymbol{\Psi}_j(\tau)$ is generated as follows:

$$\boldsymbol{\Psi}_\tau(L):=\mathbf{I}_m-\sum_{j=1}^{\infty}\boldsymbol{\Psi}_j(\tau)L^j=\mathbf{D}_\tau^{-1}(L)\mathbf{C}_\tau(L),$$

where $\mathbf{C}_\tau(L):=\mathbf{C}_1(\tau)L+\cdots+\mathbf{C}_p(\tau)L^p$ and $\mathbf{D}_\tau(L):=\mathbf{I}_m-\mathbf{D}_1(\tau)L-\cdots-\mathbf{D}_q(\tau)L^q$.

Consequently, we can present the following proposition.

Proposition 2.4. *Suppose that there is a compact set*

$$\boldsymbol{\Theta}=\{\boldsymbol{\vartheta}=[\mathrm{vec}(\mathbf{c}_0,\mathbf{C}_1,\dots,\mathbf{C}_p,\mathbf{D}_1,\dots,\mathbf{D}_q);\mathrm{vechl}(\boldsymbol{\Omega})]\mid\boldsymbol{\vartheta}\in\mathbb{R}^d\}$$

such that (1). for $\forall\tau\in[0,1]$, $\theta(\tau)$ *lies in the interior of* $\boldsymbol{\Theta}$, *(2).* $\boldsymbol{\Omega}>0$ *and* $\|\boldsymbol{\Omega}^{1/2}\varepsilon_t\|_r^2\sum_{j=1}^{\infty}|\boldsymbol{\Psi}_j|<1$ *for some* $r>6$, *(3). all the roots of* $|\mathbf{I}_m-\sum_{j=1}^{p}\mathbf{C}_j-\sum_{j=1}^{q}\mathbf{D}_j|$ *are outside the unit circle with* $\mathbf{C}_j$*'s and* $\mathbf{D}_j$*'s being squared matrices of nonnegative elements, (4).* $\mathbf{c}_0$ *is a vector of positive elements, (5).* $\mathbf{C}(L)$ *and* $\mathbf{D}(L)$ *are coprime and the formulation of the GARCH part is minimal, where* $\mathbf{C}(L):=\mathbf{C}_1L+\cdots+\mathbf{C}_pL^p$ *and* $\mathbf{D}(L):=\mathbf{I}_m-\mathbf{D}_1L-\cdots-\mathbf{D}_qL^q$. *Then the results Theorems 2.1 and 2.2 hold for the model (2.9).*

For the identification conditions of GARCH processes, we refer readers to Proposition 3.4 of Jeantheau (1998) that shows the minimal representation is enough to ensue Assumption 3 holds.

Example 3. Consider a time-varying VARMA-GARCH model

$$\begin{aligned}\mathbf{x}_t&=\mathbf{a}(\tau_t)+\sum_{j=1}^{p}\mathbf{A}_j(\tau_t)\mathbf{x}_{t-j}+\mathbf{v}_t+\sum_{j=1}^{q}\mathbf{B}_j(\tau_t)\mathbf{v}_{t-j},\quad\mathbf{v}_t=\mathrm{diag}\left(h_{1,t}^{1/2},\dots,h_{m,t}^{1/2}\right)\boldsymbol{\eta}_t,\quad t=1,\dots,T,\\ \mathbf{h}_t&=\mathbf{c}_0(\tau_t)+\sum_{j=1}^{p'}\mathbf{C}_j(\tau_t)\left(\mathbf{v}_{t-j}\odot\mathbf{v}_{t-j}\right)+\sum_{j=1}^{q'}\mathbf{D}_j(\tau_t)\mathbf{h}_{t-j},\quad\boldsymbol{\eta}_t=\boldsymbol{\Omega}^{1/2}(\tau_t)\varepsilon_t,\end{aligned} \tag{2.13}$$

where $h_{j,t}$ stands for the jth element of $\mathbf{h}_t$ and

$$\boldsymbol{\Omega}(\tau)=\begin{bmatrix} 1 & \rho_{1,2}(\tau) & \cdots & \rho_{1,m}(\tau) \\ \rho_{1,2}(\tau) & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{m-1,m}(\tau) \\ \rho_{1,m}(\tau) & \rho_{m-1,m}(\tau) & \ddots & 1 \end{bmatrix}.$$

For $\forall\tau\in[0,1]$, we obtain that

$$\widetilde{\mathbf{x}}_t(\tau)=\boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{t-1}(\tau),\widetilde{\mathbf{x}}_{t-2}(\tau),\dots;\theta(\tau)\right)+\widetilde{\mathbf{v}}_t(\tau), \tag{2.14}$$

where $\boldsymbol{\mu}\left(\widetilde{\mathbf{x}}_{t-1}(\tau),\widetilde{\mathbf{x}}_{t-2}(\tau),\dots;\theta(\tau)\right)=\mathbf{B}_\tau^{-1}(1)\mathbf{a}(\tau)+\sum_{j=1}^{\infty}\boldsymbol{\Gamma}_j(\tau)\widetilde{\mathbf{x}}_{t-j}(\tau)$,

$$\begin{aligned}\theta(\tau)&=\left[\mathrm{vec}(\mathbf{a}(\tau),\mathbf{A}_1(\tau),\dots,\mathbf{A}_p(\tau),\mathbf{B}_1(\tau),\dots,\mathbf{B}_q(\tau),\mathbf{c}_0(\tau),\mathbf{C}_1(\tau),\dots,\mathbf{C}_{p'}(\tau),\right.\\ &\qquad\left.\mathbf{D}_1(\tau),\dots,\mathbf{D}_{q'}(\tau));\mathrm{vechl}(\boldsymbol{\Omega}(\tau))\right],\\ &\qquad\mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau),\widetilde{\mathbf{x}}_{t-2}(\tau),\dots;\theta(\tau)\right),\\ &=\mathrm{diag}^{1/2}\left(\mathbf{D}_\tau^{-1}(1)\mathbf{c}_0(\tau)+\sum_{j=1}^{\infty}\boldsymbol{\Psi}_j(\tau)\left(\widetilde{\mathbf{v}}_{t-j}(\tau)\odot\widetilde{\mathbf{v}}_{t-j}(\tau)\right)\right)\boldsymbol{\Omega}^{1/2}(\tau),\\ \widetilde{\mathbf{v}}_t(\tau)&=\mathbf{H}\left(\widetilde{\mathbf{x}}_{t-1}(\tau),\widetilde{\mathbf{x}}_{t-2}(\tau),\dots;\theta(\tau)\right)\varepsilon_t.\end{aligned} \tag{2.15}$$

Note that $\boldsymbol{\Gamma}_j(\tau)$ and $\boldsymbol{\Psi}_j(\tau)$ are generated as follows:

$$\mathbf{I}_m - \sum_{j=1}^{\infty} \boldsymbol{\Gamma}_j(\tau)L^j := \mathbf{B}_\tau^{-1}(L)\mathbf{A}_\tau(L),$$

$$\boldsymbol{\Psi}_\tau(L) := \mathbf{I}_m - \sum_{j=1}^{\infty} \boldsymbol{\Psi}_j(\tau)L^j = \mathbf{D}_\tau^{-1}(L)\mathbf{C}_\tau(L), \tag{2.16}$$

where $\mathbf{A}_\tau(L) := \mathbf{I}_m - \mathbf{A}_1(\tau)L - \cdots - \mathbf{A}_p(\tau)L^p$, $\mathbf{B}_\tau(L) := \mathbf{I}_m + \mathbf{B}_1(\tau)L + \cdots + \mathbf{B}_q(\tau)L^q$ $\mathbf{C}_\tau(L) := \mathbf{C}_1(\tau)L + \cdots + \mathbf{C}_{p'}(\tau)L^{p'}$ and $\mathbf{D}_\tau(L) := \mathbf{I}_m - \mathbf{D}_1(\tau)L - \cdots - \mathbf{D}_{q'}(\tau)L^{q'}$.

We then have the following proposition.

Proposition 2.5. *Suppose that there is a compact set*

$$\boldsymbol{\Theta} = \{\boldsymbol{\vartheta} = [\text{vec}(\mathbf{a}, \mathbf{A}_1, \dots, \mathbf{A}_p, \mathbf{B}_1, \dots, \mathbf{B}_q, \mathbf{c}_0, \mathbf{C}_1, \dots, \mathbf{C}_{p'}, \mathbf{D}_1, \dots, \mathbf{D}_{q'}); \text{vechl}(\boldsymbol{\Omega})] \mid \boldsymbol{\vartheta} \in \mathbb{R}^d\}$$

such that (1). for $\forall\tau \in [0,1]$, $\theta(\tau)$ *lies in the interior of* $\boldsymbol{\Theta}$, *(2).* $\det(\mathbf{A}(L)\mathbf{B}(L)) \neq 0$ *for all* $|L| \le 1$, *where* $\mathbf{A}(L) := \mathbf{I}_m - \mathbf{A}_1 L - \cdots - \mathbf{A}_p L^p$ *and* $\mathbf{B}(L) := \mathbf{I}_m + \mathbf{B}_1 L + \cdots + \mathbf{B}_q L^q$ *are coprime and satisfy some necessary identification conditions, (3).* $\boldsymbol{\Omega} > 0$ *and* $\|\boldsymbol{\Omega}^{1/2}\varepsilon_t\|_r^2 \sum_{j=1}^{\infty} |\boldsymbol{\Psi}_j| < 1$ *for some* $r > 6$, *(4). all the roots of* $|\mathbf{I}_m - \sum_{j=1}^{p'} \mathbf{C}_j - \sum_{j=1}^{q'} \mathbf{D}_j|$ *are outside the unit circle with* $\mathbf{C}_j$*'s and* $\mathbf{D}_j$*'s being squared matrices of nonnegative elements, (5).* $\mathbf{c}_0$ *is a vector of positive elements, (6).* $\mathbf{C}(L)$ *and* $\mathbf{D}(L)$ *are coprime and the formulation of the GARCH part is minimal, where* $\mathbf{C}(L) := \mathbf{C}_1 L + \cdots + \mathbf{C}_{p'} L^{p'}$ *and* $\mathbf{D}(L) := \mathbf{I}_m - \mathbf{D}_1 L - \cdots - \mathbf{D}_{q'} L^{q'}$. *Then the results Theorems 2.1 and 2.2 hold for the model (2.13).*

In the following section, we conduct numerical studies using both simulated and real data to evaluate the finite-sample performance of the proposed estimation and inferential methods.

3. Simulation studies

In this section, we first present the details of the numerical implementations in Section 3.1, and then conduct extensive simulations in Section 3.2.

3.1. Numerical implementation

Throughout the numerical studies, the Epanechnikov kernel $K(u) = 0.75(1-u^2)I(|u| \le 1)$ is adopted. Following Zhou and Wu (2010), we use $\widetilde{h} = 2\widehat{h}$ for the biased corrected estimator, where $\widehat{h}$ is the bandwidth selected by the cross-validation method of Richter and Dahlhaus (2019).

Specifically, define the leave-one-out local linear QMLE

$$(\widehat{\theta}_{h,-t}(\tau), h\widehat{\theta}_{h,-t}^{(1)}(\tau)) = \underset{(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)\in\mathbf{E}_T(r)}{\text{argmax}}\ \mathcal{L}_{T,-t}^c(\tau, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2), \tag{3.1}$$

where

$$\mathcal{L}_{T,-t}^c(\tau, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2) = \frac{1}{T}\sum_{s=1,\neq t}^{T} \ell(\mathbf{x}_s, \mathbf{z}_{s-1}^c; \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \cdot (\tau_s - \tau)/h) K_h(\tau_s - \tau).$$

Then, the bandwidth is chosen by

$$\widehat{h} = \underset{h}{\text{argmax}}\, T^{-1}\sum_{t=1}^{T} \ell(\mathbf{x}_t, \mathbf{z}_{t-1}^c; \widehat{\theta}_{h,-t}(\tau_t)). \tag{3.2}$$

As shown in Richter and Dahlhaus (2019), the cross validation method works well as long as $\nabla\ell$ is uncorrelated, which implies that the desirable property should hold in our case. It should be pointed out that the cross-validation bandwidth is optimal for parameter estimation in the sense of mean squared errors, which may not be optimal for constructing uniform confidence bands. However, it is unclear how to select the optimal bandwidth for uniform inference and our simulation results below show that the behavior of the estimated UCB is not sensitive to the choices of bandwidths. For simplicity, we can use the estimation-based optimal bandwidth when constructing UCB.

Notably, when considering some specific models, the implementation may be further simplified. We provide more discussions along this line in Appendix B.4.

Table 1
Empirical coverage probabilities of the UCB for DGP 1.

	$\widetilde{h}$	$\alpha_1(\cdot)$	$\alpha_2(\cdot)$	$\mathbf{B}_1(\cdot)$	$\boldsymbol{\Omega}(\cdot)$
$T = 500$	0.35	0.905	0.915	0.889	0.905
	0.4	0.912	0.930	0.897	0.909
	0.45	0.915	0.930	0.898	0.919
	0.5	0.905	0.945	0.901	0.920
$T = 1000$	0.3	0.960	0.960	0.947	0.947
	0.35	0.940	0.967	0.940	0.930
	0.4	0.947	0.959	0.948	0.939
	0.45	0.957	0.957	0.947	0.937

3.2. Simulation results

In the simulation studies, we examine the empirical coverage probabilities of UCB for the 95% nominal level. We consider the time-varying VARMA(2,1) and time-varying VAR(1)-GARCH(1,1) model as follows:

1. DGP 1 : $\mathbf{x}_t = a_1(\tau_t)\mathbf{x}_{t-1} + a_2(\tau_t)\mathbf{x}_{t-2} + \boldsymbol{\eta}_t + \mathbf{B}_1(\tau_t)\boldsymbol{\eta}_{t-1}$, $\boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau_t)\boldsymbol{\epsilon}_t$, where $\{\boldsymbol{\epsilon}_t\}$ are i.i.d. draws from $N(\mathbf{0}_{2\times 1}, \mathbf{I}_2)$, $a_1(\tau) = 0.6\exp(\tau - 1)$, $a_2(\tau) = -0.3\exp(\tau - 1)$,
$$\mathbf{B}_1(\tau) = \begin{bmatrix} 0.5\exp\{\tau - 0.5\} & -0.8(\tau-0.5)^2 \\ -0.8(\tau-0.5)^2 & 0.5 + 0.3\sin(\pi\tau) \end{bmatrix},$$
$$\boldsymbol{\omega}(\tau) = \begin{bmatrix} 1.5 + 0.2\exp\{0.5-\tau\} & 0 \\ 0.2\exp\{0.5-\tau\} & 1.5 + 0.5(\tau-0.5)^2 \end{bmatrix}.$$
Here we use the final equations form to ensure the uniqueness of the VARMA representation, while the order of VARMA is set to be $(2, 1)$. The choices of $a_1(\tau)$, $a_2(\tau)$ and $\mathbf{B}_1(\tau)$ correspond to moderate persistence in the VARMA dynamics. These settings are common in the VARMA literature and are consistent with some stylized facts of macro variables (cf., Athanasopoulos and Vahid (2008)).
2. DGP 2 : $\mathbf{x}_t = \mathbf{A}_1(\tau_t)\mathbf{x}_{t-1} + \mathbf{v}_t$ and $\mathbf{v}_t = \mathrm{diag}(h_{1,t}^{1/2}, \dots, h_{m,t}^{1/2})\boldsymbol{\eta}_t$, where $\boldsymbol{\eta}_t = \boldsymbol{\Omega}^{1/2}(\tau_t)\boldsymbol{\epsilon}_t$, $\mathbf{h}_t = \mathbf{c}_0(\tau_t) + \mathbf{C}_1(\tau_t)\left(\mathbf{v}_{t-1}\odot\mathbf{v}_{t-1}\right) + \mathbf{D}_1(\tau_t)\mathbf{h}_{t-1}$, $\{\boldsymbol{\epsilon}_t\}$ are i.i.d. draws from $N(\mathbf{0}_{2\times 1}, \mathbf{I}_2)$, $\mathbf{c}_0(\tau) = [2\exp\{0.5\tau - 0.5\}, 3 + 0.2\cos(\tau)]^\top$,
$$\mathbf{A}_1(\tau) = \begin{bmatrix} 0.5\exp\{\tau-0.5\} & -0.2\exp\{\tau-1\} \\ -0.2\cos(\pi\tau) & 0.6\exp\{-\tau-0.5\} \end{bmatrix},$$
$$\mathbf{C}_1(\tau) = \begin{bmatrix} 0.4 + 0.05\cos(\tau) & 0.05(\tau-0.5)^2 \\ 0.05(\tau-0.5)^2 & 0.4 + 0.05\sin(\tau) \end{bmatrix},$$
$$\mathbf{D}_1(\tau) = \begin{bmatrix} 0.4 - 0.1\cos(\tau) & 0 \\ 0 & 0.3 - 0.1\sin(\tau) \end{bmatrix},$$
$$\boldsymbol{\Omega}(\tau) = \begin{bmatrix} 1 & 0.3\sin(\tau) \\ 0.3\sin(\tau) & 1 \end{bmatrix}.$$
This model specification is the same as the one that we used in our empirical section. We set the order of GARCH process to be $(1, 1)$ since GARCH$(1, 1)$ models are typically used in practice and higher order GARCH models are unnecessary (cf., Andreou and Werker (2015)). In addition, the time-varying functions $\mathbf{C}_1(\tau)$, $\mathbf{D}_1(\tau)$ and $\boldsymbol{\Omega}(\tau)$ are set to be roughly the same as our estimates for the stocks returns we analyze in the next section.

Let the sample size be $T \in \{500, 1000\}$ ($T \in \{1000, 2000, 4000\}$) for the VARMA model (the VAR-GARCH model). We conduct 1000 replications for each choice of T. Several different bandwidths close to $\widetilde{h}$ are reported to check the sensitivity of bandwidth selection.

We present the empirical coverage probabilities associated with the UCB in Tables 1–2. For the vector- or matrix-valued unknown coefficients, we take an average across the elements. A few facts emerge from the tables. First, the finite sample coverage probabilities are smaller than their nominal level when $T = 500$ ($T = 1000, 2000$) for the VARMA model (the GARCH/VAR-GARCH model), but are fairly close to their nominal level as $T = 1000$ ($T = 4000$) for the VARMA model (the VAR-GARCH model). Second, the behavior of the estimated uniform confidence bands is not sensitive to the choices of bandwidths. The leave-one-out cross-validation method tends to select a larger bandwidth for the conditional variance model compared to that of the conditional mean model. Third, the conditional variance model requires more data than the conditional mean model to achieve a reasonable finite sample performance. This point is also illustrated by the simulation results of GARCH models, which are reported in Appendix B.5 due to space constraints.

4. A real data example

In this section, we investigate both temporal and contemporaneous connectedness between the Chinese and U.S. stock markets using a time-varying VAR-GARCH model. Our framework facilitates the study of connectedness in both temporal and contemporaneous levels, as well as time-variations in these spillovers. These are of great importance in practice.

Table 2
Empirical coverage probabilities of the UCB for DGP 2.

	$\widetilde{h}$	$\mathbf{A}_1(\cdot)$	$\mathbf{c}_0(\cdot)$	$\mathbf{C}_1(\cdot)$	$\mathbf{D}_1(\cdot)$	$\boldsymbol{\Omega}(\cdot)$
$T = 1000$	0.60	0.953	0.833	0.829	0.751	0.947
	0.65	0.945	0.841	0.837	0.749	0.940
	0.70	0.955	0.864	0.847	0.787	0.950
	0.75	0.949	0.847	0.854	0.779	0.945
$T = 2000$	0.55	0.968	0.905	0.866	0.852	0.922
	0.60	0.957	0.883	0.878	0.847	0.927
	0.65	0.955	0.889	0.884	0.857	0.912
	0.70	0.965	0.907	0.891	0.902	0.950
$T = 4000$	0.50	0.963	0.920	0.903	0.897	0.940
	0.55	0.967	0.907	0.917	0.904	0.925
	0.60	0.960	0.917	0.934	0.927	0.929
	0.65	0.963	0.942	0.918	0.914	0.959

4.1. A brief literature review

Researchers recently have applied several statistical measures of association (e.g., Granger causality) to asset returns to analyze the spillovers between financial agents (e.g., Hong, 2001; Diebold and Yilmaz, 2009, 2014; Geraci and Gnabo, 2018). For example, Geraci and Gnabo (2018) apply a Bayesian time-varying VAR(1) model to estimate the dynamic network of return spillover effects at U.S. institutional levels, in which the time-varying cross-autoregressive coefficients are interpreted as the strength of temporal and directed connectedness between financial institutions over time.

In addition, there is a growing literature to study the relationship of Chinese and U.S. stock markets (e.g., Zhang and Li, 2014; Pan et al., 2022), as the Chinese stock market has become the world's second largest stock market after 2009. Understanding the interactions among different financial markets is important for investors and policymakers (BenSaïda, 2019). For instance, high equity market interdependence implies poor diversification benefits from portfolios, but highlights the possibility of better hedging benefits. Previous research documents a strong positive link between the degree of globalization and equity market interdependence (Baele, 2005). Along this line of research, one important question is that whether the interdependence between Chinese and U.S. stock markets has increased over time due to globalization so that estimates from historical data are unreliable for modern policy analysis, asset pricing and risk management. The existing results present many discrepancies, which may be due to the fact that the relationship evolves with time. Apparently, the results also indicate that one should use a time-varying VAR-GARCH model to accommodate potential nonstationarity inherited in these financial variables. In what follows, we address these issues using the newly proposed approach.

4.2. Empirical analysis

To study time-varying temporal and contemporaneous connectedness simultaneously, we consider the following time-varying VAR(1)-GARCH(1,1) model:

$$\begin{aligned}\mathbf{x}_t &= \mathbf{a}(\tau_t) + \mathbf{A}(\tau_t)\mathbf{x}_{t-1} + \mathbf{v}_t, \quad \mathbf{v}_t = \mathrm{diag}(h_{1,t}^{1/2}, h_{2,t}^{1/2})\boldsymbol{\eta}_t, \\ \boldsymbol{\eta}_t &= \boldsymbol{\Omega}^{1/2}(\tau_t)\varepsilon_t, \quad \mathbf{h}_t = \mathbf{c}(\tau_t) + \mathbf{C}(\tau_t)\left(\mathbf{v}_{t-1} \odot \mathbf{v}_{t-1}\right) + \mathbf{D}(\tau_t)\mathbf{h}_{t-1},\end{aligned} \tag{4.1}$$

where $\mathbf{x}_t$ contains Chinese and U.S. index log-returns, which are calculated based on weekly Shanghai Stock Exchange (SSE) Composite Index and S&P 500 Index as they are the most comprehensive and diversified stock indices.[2] Note that the cross-autoregressive coefficient $A_{ij}(\tau_t)$ represents the strength of the temporal (directed) spillover from individual j to i at period t, while the off-diagonal elements of time-varying correlation matrix $\boldsymbol{\Omega}(\tau)$ capture time-varying conditional correlations between errors, which can be interpreted as the contemporaneous (undirected) spillovers occurring between the stock markets.

The estimation is conducted in exactly the same way as in Section 3.1. Here, we set the lag lengths of both VAR and GARCH settings to be 1 following Andreou and Werker (2015), in which they develop formal specification tests for univariate AR-GARCH models and find that AR(1)-GARCH(1,1) process is enough to capture the dynamics of stock market weekly returns. The sample employed in this study spanning from July 1997 to December 2021 provides 1239 observations.[3] Fig. 1 plots the two weekly returns as well as sample autocorrelation functions of squared data, which shows the typical "volatility clustering" phenomenon.

We now fit the data to a time-varying VAR(1)-GARCH(1,1) model and are particularly interested in whether the temporal and contemporaneous spillovers vary over time. We first consider the time-varying contemporaneous correlations. Fig. 2 plots the

[2] Weekly returns are commonly used in the literature on return spillover effects (e.g., Baele, 2005; Beirne et al., 2010) to avoid the non-synchronicity of daily data. The non-synchronicity of daily data arises from the fact that the national stock exchanges are subject to different national, religious, and other holidays, unexpected events, and other occasions. In addition, we only use the available data, removing the missing period since this method is desirable particularly when a missing period is just a few time points (cf., Park (2004), p. 661).

[3] The data are collected from Yahoo Finance at https://finance.yahoo.com/.

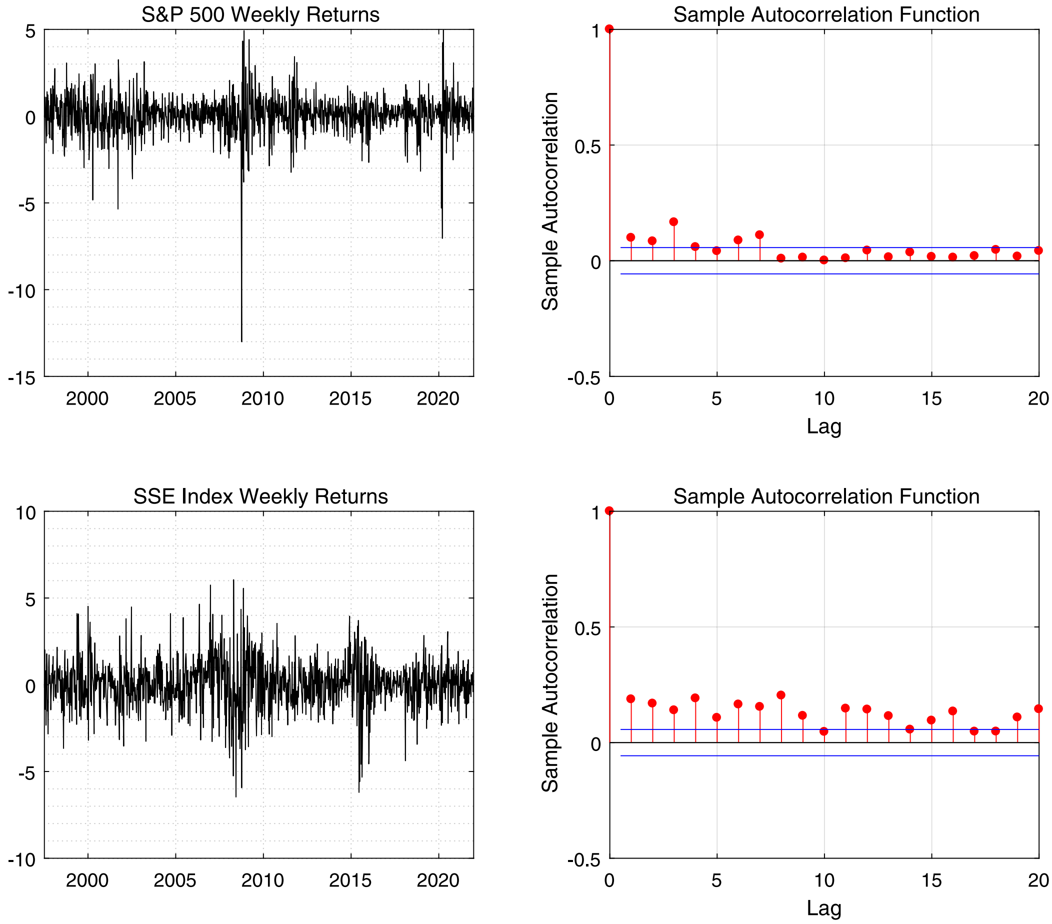

Fig. 1. S&P 500 and SSE Index returns as well as sample autocorrelation functions of squared data.

estimates of time-varying conditional correlations (black line), as well as the 95% uniform confidence bands (red dot-dashed line), the 95% pointwise confidence intervals (black dashed line), the estimate of the parameter assuming constancy (blue line) and the estimates of the dynamic conditional correlations (DCC) using a VAR-DCC-GARCH model (black dot line). Based on the uniform confidence bands and the estimates obtained from constant VAR-GARCH model, we can conclude that the conditional correlations vary with respect to time at the 5% significance level. By examining the point-wise confidence intervals, we can conclude that the two stock markets are not significantly correlated before 2005, but this relationship has been greatly enhanced in recent years. Interestingly, as clearly presented in Fig. 2, the contemporaneous interdependence between the two stock markets is increasing over time regardless of the huge impact of major events such as the trade friction and COVID-19 pandemic. This result has important implications for investment and risk management. For example, it implies that Chinese and U.S. investors who use cross-country portfolio strategies to eliminate country specific risks may be benefit from hedging. Meanwhile, the hedge ratio should be adjusted to account for the most recent information since the correlations between Chinese and U.S. stock markets are time-varying.

For comparison with the popular DCC-GARCH models, we also plot the estimates of dynamic conditional correlations obtained from a VAR-DCC-GARCH model in Fig. 2. It can be seen that the DCC estimates also exhibit an upward trend, roughly fluctuate around our smooth time-varying estimates, and almost lie within the estimated UCB of time-varying conditional correlations. However, the DCC estimates are quite volatile, and it is unclear why the DCC estimates decline persistently from 2012 to 2015 given the fact that the DCC estimates still exhibit an upward trend after 2015 regardless of the impact of the trade friction and COVID-19 pandemic. It should be noted that DCC represents the dynamic conditional covariances of the standardized residuals, and thus does not yield dynamic conditional correlations. Also, DCC yields inconsistent two step estimators (Caporin and McAleer, 2013).

We next investigate the time-varying temporal spillovers. Fig. 3 plots the estimates of time-varying cross-autoregressive coefficients, as well as the 95% uniform confidence bands (red dot-dashed line), the 95% pointwise confidence intervals (black dashed line) and the estimates of the parameters assuming constancy (blue line). The uniform confidence bands show that the temporal return spillovers from the U.S. stock market to the Chinese stock market significantly vary over time, while the temporal

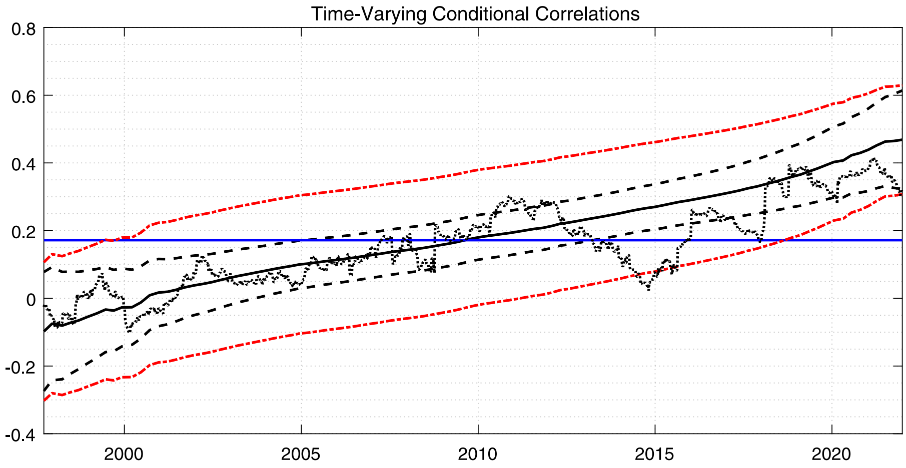

Fig. 2. Time-varying conditional correlations between the Chinese and U.S. stock markets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

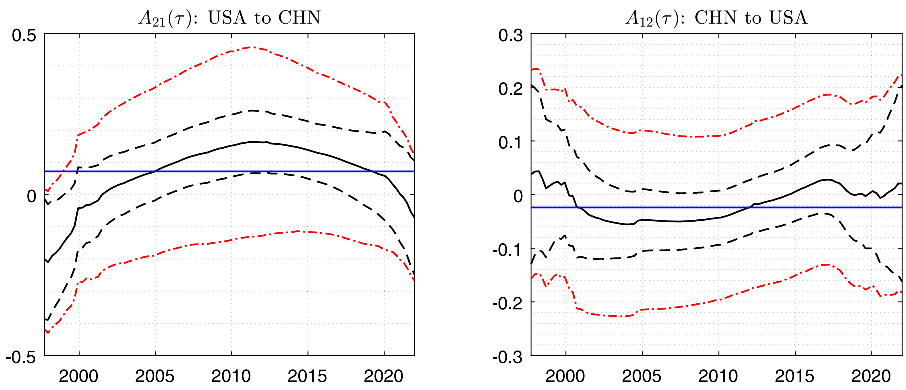

Fig. 3. Time-varying cross-autoregressive coefficients. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

return spillovers from the Chinese stock market to the U.S. stock market is not significantly from zero. By examining the point-wise confidence intervals, we find that the temporal return spillovers from the U.S. stock market to the Chinese stock market are not significant before 2005. However, during some specific time periods from 2006–2018, the point-wise intervals display locally significant return spillovers from the U.S. stock market to the Chinese stock market and this return spillover effect declines gradually after 2015. These results are consistent with the consensus that there exist significant return spillovers from mature markets to emerging markets but not vice versa (e.g., Beirne et al., 2010).

Finally, we plot the estimates of time-varying GARCH coefficients in Fig. 4 to examine whether the second-order dynamics of these two market returns vary over time. From this figure, we can see that the second-order dynamics of U.S. stock market is identified as time-varying since the uniform confidence bands do not contain the estimates of the parameter assuming constancy, while the dynamics of Chinese stock market seem to be time-invariant. Interestingly, we find that the dynamics of U.S. stock market change dramatically in 2001, which may be due to "the Internet bubble bursting" in stock market, while during this period Chinese stock market is still isolated from the rest of the world. To provide more empirical evidence, we also evaluate the conditional volatility implied by our time-varying model and its parametric counterpart. Fig. 5 plots the estimates of conditional volatility (multiplied by 2) obtained from our time-varying model (red line) and its parametric counterpart (blue line), as well as the absolute value of stock returns (black line). From Fig. 5, it can be seen that the estimates of conditional volatility from these two models are roughly the same except for some specific periods. Specifically, for S&P 500 returns, the estimated conditional volatility obtained from our time-varying is more persistent (or smoother) than that obtained from a constant model during 1999–2003. This result is consistent with Gallo and Otranto (2012), who find that the realized volatility of S&P 500 returns has regime-specific behavior and its persistence has declined after the period 1998–2003. In addition, for SSE index returns, the estimated conditional volatility obtained from our time-varying model is smaller than that obtained from a constant model after 2015 when the stock market is quite calm or of low volatility, but these two estimates are roughly the same when the stock markets are of high volatility.

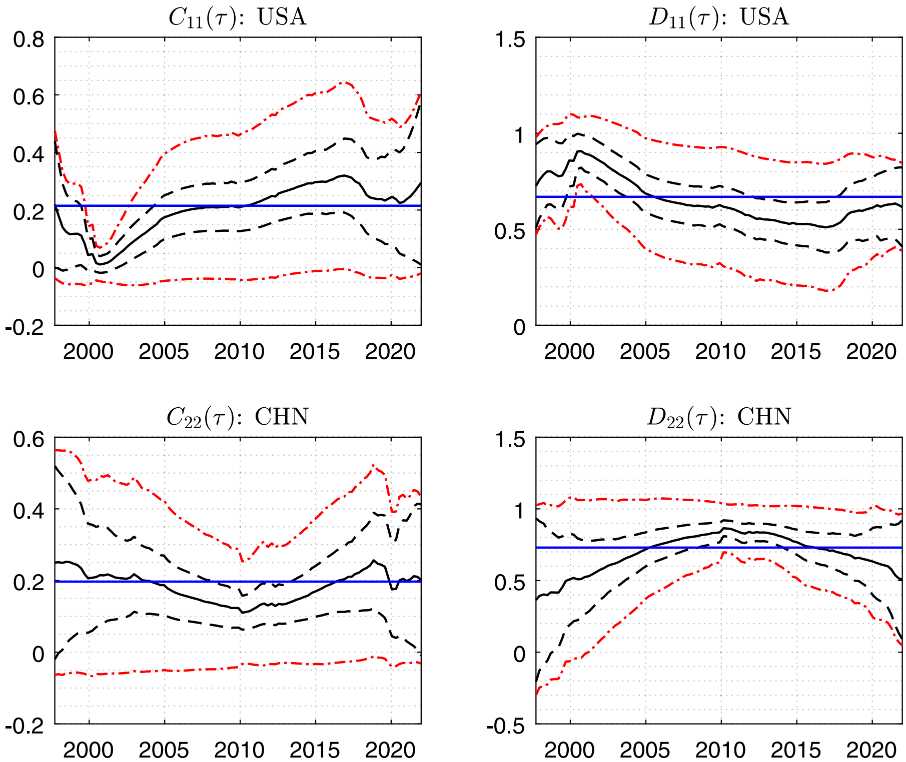

Fig. 4. Time-varying GARCH coefficients. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Conclusions

In this paper, we consider a wide class of time-varying multivariate causal processes which nests many classic and new examples as special cases. We first prove the existence of a weakly dependent stationary approximation for the model (1.1) which is the foundation to establish the corresponding asymptotic properties. Afterwards, we consider the QMLE estimation, and provide both point-wise and uniform inferences on the coefficient functions. In addition, we demonstrate the theoretical findings through both simulated and real data examples. In particular, we show the empirical relevance of our study using an application to evaluate the temporal return/volatility spillovers and conditional correlations between the stock markets of China and U.S. We find that (1) the contemporaneous interdependence between the two stock markets is statistically significant and is increasing over time and (2) the temporal return spillovers from the U.S. stock market to the Chinese stock market significantly vary over time and can be only detected at some local times.

There are several directions for possible extensions. The first one is to consider quantile regression methods for such locally stationary multivariate causal processes. The second one is to propose a more powerful L_2 test based on the weighted integrated squared errors for testing whether some coefficients are time-invariant. We wish to leave such issues for future study.

Acknowledgments

The authors would like to thank the Co-Editor, the Associate Editor and two referees for their valuable comments, which greatly improved earlier versions of this paper. Thanks also go to George Athanasopoulos, David Frazier and Gael Martin for their constructive comments on earlier versions of this paper. The authors also acknowledge the useful comments by the participants of several seminars. Gao and Peng would also like to acknowledge the Australian Research Council Discovery Projects Program for its financial support under Grant Numbers: DP200102769 & DP210100476. Yan acknowledges the financial support by the National Natural Science Foundation of China (Grant Number: 72303142) and Fundamental Research Funds for the Central Universities, China (Grant Numbers: 2022110877 & 2023110099).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2024.105671.

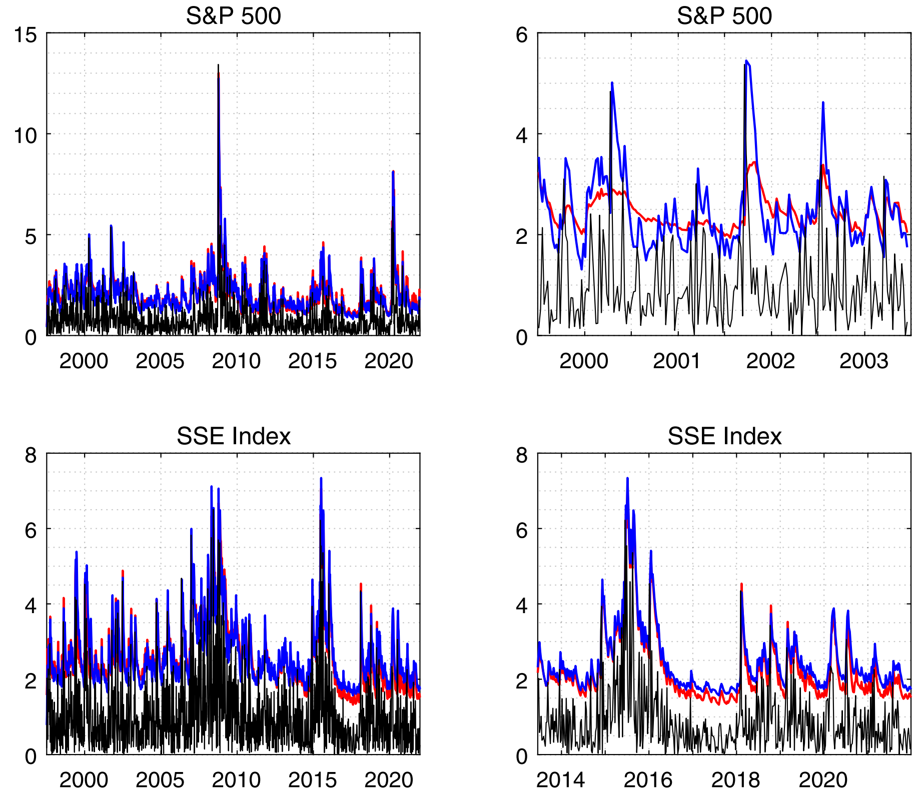

Fig. 5. Absolute stock returns and the estimated conditional volatility. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

References

Andreou, E., Werker, B.J., 2015. Residual-based rank specification tests for AR-GARCH type models. J. Econometrics 185 (2), 305–331.
Athanasopoulos, G., Vahid, F., 2008. VARMA versus VAR for macroeconomic forecasting. J. Bus. Econom. Statist. 26 (2), 237–252.
Augustyniak, M., Badescu, A., Bégin, J.-F., 2023. A discrete-time hedging framework with multiple factors and fat tails: On what matters. J. Econometrics 232 (2), 416–444.
Baele, L., 2005. Volatility spillover effects in European equity markets. J. Financ. Quant. Anal. 40 (2), 373–401.
Bai, J., 1997. Estimation of a change point in multiple regression models. Rev. Econ. Stat. 79 (4), 551–563.
Bardet, J.-M., Wintenberger, O., 2009. Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. Ann. Statist. 37 (5B), 2730–2759.
Bauwens, L., Laurent, S., Rombouts, J.V., 2006. Multivariate GARCH models: a survey. J. Appl. Econometrics 21 (1), 79–109.
Beirne, J., Caporale, G.M., Schulze-Ghattas, M., Spagnolo, N., 2010. Global and regional spillovers in emerging stock markets: A multivariate GARCH-in-mean analysis. Emerg. Mark. Rev 11 (3), 250–260.
BenSaïda, A., 2019. Good and bad volatility spillovers: An asymmetric connectedness. J. Financial Mark. 43, 78–95.
Berger, T., Morley, J., Wong, B., 2023. Nowcasting the output gap. J. Econometrics 232 (1), 18–34.
Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2020. Multivariate leverage effects and realized semicovariance GARCH models. J. Econometrics 217 (2), 411–430.
Caporin, M., McAleer, M., 2013. Ten things you should know about the dynamic conditional correlation representation. Econometrics 1 (1), 115–126.
Chen, X., Christensen, T.M., 2018. Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. Quant. Econom. 9 (1), 39–84.
Chen, B., Hong, Y., 2012. Testing for smooth structural changes in time series models via nonparametric regression. Econometrica 80 (3), 1157–1183.
Chen, L., Wang, W., Wu, W.B., 2022. Inference of breakpoints in high-dimensional time series. J. Amer. Statist. Assoc. 540 (117), 1951–1963.
Dahlhaus, R., 1996. On the Kullback–Leibler information divergence of locally stationary processes. Stochastic Process. Appl. 62 (1), 139–168.
Dahlhaus, R., 1997. Fitting time series models to nonstationary processes. Ann. Statist. 25 (1), 1–37.
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. Ann. Statist. 34 (3), 1075–1114.
Diebold, F.X., Yilmaz, K., 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. Econ. J. 119 (534), 158–171.
Diebold, F.X., Yilmaz, K., 2014. On the network topology of variance decompositions: Measuring the connectedness of financial firms. J. Econometrics 182 (1), 119–134.
Doukhan, P., 2012. Mixing: Properties and Examples, vol. 85, Springer Science & Business Media.
Engle, R.F., Siriwardane, E.N., 2018. Structural GARCH: the volatility-leverage connection. Rev. Financ. Stud. 31 (2), 449–492.
Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Springer.
Gallo, G.M., Otranto, E., 2012. Realized Volatility and Change of Regimes. Technical Report.
Geraci, M.V., Gnabo, J.-Y., 2018. Measuring interconnectedness between financial institutions with Bayesian time-varying vector autoregressions. J. Financ. Quant. Anal. 53 (3), 1371–1390.

Hall, P., 1980. Estimating probabilities for normal extremes. Adv. Appl. Probab. 12 (2), 491–500.
Hong, Y., 2001. A test for volatility spillover with application to exchange rates. J. Econometrics 103 (1–2), 183–224.
Inoue, A., Kilian, L., 2022. Joint Bayesian inference about impulse responses in VAR models. J. Econometrics 231 (2), 457–476.
Jeantheau, T., 1998. Strong consistency of estimators for multivariate ARCH models. Econom. Theory 14 (1), 70–86.
Karmakar, S., Richter, S., Wu, W.B., 2022. Simultaneous inference for time-varying models. J. Econometrics 227 (2), 408–428.
Kilian, L., Lütkepohl, H., 2017. Structural Vector Autoregressive Analysis. Cambridge University Press.
Li, J., Liao, Z., 2020. Uniform nonparametric inference for time series. J. Econometrics 219 (1), 38–51.
Li, J., Liao, Z., Zhou, W., 2023. Uniform nonparametric inference for spatially dependent panel data. J. Bus. Econom. Statist. 1–22, in press.
Li, Q., Racine, J.S., 2007. Nonparametric Econometrics: Theory and Practice. Princeton University Press.
Ling, S., 2003. Adaptive estimators and tests of stationary and nonstationary short-and long-memory ARFIMA–GARCH models. J. Amer. Statist. Assoc. 98 (464), 955–967.
Ling, S., McAleer, M., 2003. Asymptotic theory for a vector ARMA-GARCH model. Econom. Theory 19 (2), 280–310.
Lütkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. Springer Science & Business Media.
Miranda-Agrippino, S., Ricco, G., 2021. The transmission of monetary policy shocks. Am. Econ. J.: Macroecon. 13 (3), 74–107.
Pan, Q., Mei, X., Gao, T., 2022. Modeling dynamic conditional correlations with leverage effects and volatility spillover effects: Evidence from the Chinese and US stock markets affected by the recent trade friction. North Am. J. Econom. Finance 59, 101591.
Park, N.K., 2004. A guide to using event study methods in multi-country settings. Strateg. Manage. J. 25 (7), 655–668.
Phillips, P.C.B., Li, D., Gao, J., 2017. Estimating smooth structural change in cointegration models. J. Econometrics 196 (1), 180–195.
Preuss, P., Puchstein, R., Dette, H., 2015. Detection of multiple structural breaks in multivariate time series. J. Amer. Statist. Assoc. 110 (510), 654–668.
Richter, S., Dahlhaus, R., 2019. Cross validation for locally stationary processes. Ann. Statist. 47 (4), 2145–2173.
Schorfheide, F., Song, D., 2015. Real-time forecasting with a mixed-frequency VAR. J. Bus. Econom. Statist. 33 (3), 366–380.
Stock, J.H., Watson, M.W., 2001. Vector autoregressions. J. Econ. Perspect. 15 (4), 101–115.
Tong, H., 1990. Non-Linear Time Series: A Dynamical Systems Approach. Oxford University Press.
Truquet, L., 2017. Parameter stability and semiparametric inference in time varying auto-regressive conditional heteroscedasticity models. J. R. Stat. Soc. Ser. B Stat. Methodol. 79 (5), 1391–1414.
Vogt, M., 2012. Nonparametric regression for locally stationary time series. Ann. Statist. 40 (5), 2601–2633.
Withers, C., 1981. Conditions for linear processes to be strong-mixing. Z. Wahrscheinlichkeitstheor. Verwandte Geb. 57 (4), 477–480.
Wu, W.B., 2005. Nonlinear system theory: Another look at dependence. Proc. Natl. Acad. Sci. 102 (40), 14150–14154.
Wu, W.B., Shao, X., 2004. Limit theorems for iterated random functions. J. Appl. Probab. 41 (2), 425–436.
Xia, Y., 1998. Bias-corrected confidence bands in nonparametric regression. J. R. Stat. Soc. Ser. B Stat. Methodol. 60 (4), 797–811.
Zhang, B., Li, X.-M., 2014. Has there been any change in the comovement between the Chinese and U.S. stock markets? Int. Rev. Econ. Finance 29, 525–536.
Zhang, T., Wu, W.B., 2012. Inference of time-varying regression models. Ann. Statist. 40 (3), 1376–1402.
Zhou, Z., Wu, W.B., 2010. Simultaneous inference of linear models with time varying coefficients. J. R. Stat. Soc. Ser. B Stat. Methodol. 72 (4), 513–531.