

TESTING AND MODELLING FOR THE STRUCTURAL CHANGE IN COVARIANCE MATRIX TIME SERIES WITH MULTIPLICATIVE FORM

Feiyu Jiang, Dong Li, Wai Keung Li and Ke Zhu

*Fudan University, Tsinghua University,
Education University of Hong Kong and University of Hong Kong*

Abstract: We first construct a new generalized Hausman test for detecting the structural change in a multiplicative form of covariance matrix time series model. This generalized Hausman test is asymptotically pivotal, and has nontrivial power in detecting a broad class of alternatives. Moreover, we propose a new semiparametric covariance matrix time series model. The proposed model has a time-varying long-run component that takes the structural change into account, and a BEKK-type short-run component that captures the temporal dependence. We propose a two-step estimation procedure to estimate this semiparametric model, and establish the asymptotics of the related estimators. Finally, the importance of the generalized Hausman test and the semiparametric model is illustrated by means of simulations and an application to realized covariance matrix data.

Key words and phrases: Covariance matrix time series model, profiled quasi maximum likelihood estimation, realized covariance matrix, semiparametric time series model, structural change testing.

1. Introduction

Matrix-variate observations are often encountered in statistical applications with structured information. An important example is the realized covariance (RCOV) matrix, which is calculated from intraday high-frequency returns and enhances the prediction ability for the covariance matrix of the underlying multivariate process (see, e.g., Barndorff-Nielsen and Shephard (2002, 2004) and Andersen et al. (2003)). Because modelling and forecasting a covariance matrix are fundamental in asset pricing, portfolio selection, and risk management, it is necessary to study the dynamics of the RCOV matrix.

The non-ignorable structured feature of the RCOV matrix is its positive definiteness. To preserve this feature, many covariance matrix time series models have been proposed based on the Wishart or matrix-F innovation distribu-

Corresponding author: Ke Zhu, Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong. E-mail: mazhuke@hku.hk.

tion, which generates random positive-definite matrices automatically. Related works in this direction include Gouriéroux, Jasiak and Sufana (2009), Golosnoy, Gribisch and Liesenfeld (2012), Yu, Li and Ng (2017), Opschoor et al. (2018), and many others. However, the aforementioned models all require stationarity, which may be less reliable in view of the ubiquity of structural changes in applications; see, for example, Aue and Horváth (2013) for an overview. In the presence of a structural change, conventional statistical methodologies could lead to a severe deterioration in predictions (Pesaran and Timmermann (2004)). Furthermore, sub-optimal long-memory models may be adopted to capture the spurious long-memory phenomenon (Mikosch and Stărică (2004); Stărică and Granger (2005)). In this paper, we show that RCOV matrix data can also exhibit the spurious long-memory phenomenon caused by a structural change. This indicates that existing long-memory models (Corsi (2009); Chiriac and Voev (2011)) for RCOV matrix data should be used with caution, because they overlook the possibility of spurious long memory. To circumvent the aforementioned dilemmas, a formal test for detecting structural changes in matrix time series models is needed, but has not been attempted before.

We propose a new generalized Hausman test for detecting a structural change in a multiplicative formulation of the covariance matrix time series model. The considered multiplicative testing model is a matrix generalization of the univariate volatility model in Feng (2004) and the multivariate one in Hafner and Linton (2010), both of which allow for abrupt and smooth structural changes with an unknown form, in the spirit of Robinson (1989, 1991). The generalized Hausman testing framework was first proposed by Chen and Hong (2012), and then later adopted and extended by Zhang and Wu (2012), Chen (2015), Chen and Hong (2016), and Fu and Hong (2019). To the best of our knowledge, our generalized Hausman test is the first formal attempt to detect structural changes for matrix time series models. Under suitable conditions, our test is shown to have a standard normal limiting null distribution, as well as nontrivial power in detecting a broad class of abrupt and smooth local alternatives.

Testing for structural changes is a long-standing problem in the literature. In addition to the generalized Hausman test, many other tests have been studied for regression or univariate/multivariate time series models; see, for example, the cumulative sum type tests in Brown, Durbin and Evans (1975), Hidalgo (1995), Su and Xiao (2008), Aue et al. (2009), and Shao and Zhang (2010); the supremum type tests in Andrews (1993), Davis, Huang and Yao (1995), Bai and Perron (1998), Hansen (2000), and Qu and Perron (2007); and the Lagrange multiplier tests in Lin and Teräsvirta (1994) and Amado and Teräsvirta (2008). In this

paper, we study the generalized Hausman test, because it has several appealing advantages over other tests. First, it can detect both abrupt and smooth changes without assuming a specified structural change alternative or knowing the type of change. Clearly, it is important for the test to avoid such potential model misspecification. Second, it is asymptotically pivotal, and thus easy-to-implement. Third, there is no need to trim the boundary region near the end points of the sample period, allowing the test to detect changes that happen very early or very late.

Once a structural change is indicated by our generalized Hausman test, it is important to have a dynamic model to capture this change. For this purpose, we propose a new semiparametric covariance matrix time series model. Our semiparametric model has a multiplicative form, combining a time-varying deterministic long-run component and a BEKK-type (Engle and Kroner (1995)) short-run component. Hence, we call it the time-varying BEKK (TVB) model. For this nonstationary TVB model, we propose a two-step estimation procedure. Specifically, we first consider a kernel estimator for the long-run component. Second, we estimate the unknown parameter vector of the BEKK-type short-run component using a profiled quasi maximum likelihood estimator (QMLE). This profiled QMLE is obtained by assuming that the model innovation follows the Wishart distribution, and profiles out the unobserved long-run component by using the kernel estimator in step one. The consistency of both the kernel estimator and the profiled QMLE is established under conditions allowing for both smooth and abrupt changes (at some finite points) in the long-run component, and the asymptotic normality of both estimators is derived under some smoothness conditions on the long-run component. To derive the asymptotics of the profiled QMLE, we show that the involved kernel estimation effect from step one is negligible. This requires nontrivial proof techniques, because the BEKK-type process depends on all past unobserved lagged variables in our TVB model.

The remainder of the paper is organized as follows. Section 2 presents the testing framework, proposes the test statistic, and derives the corresponding asymptotic theory. Section 3 introduces the semiparametric TVB model and studies its two-step estimation procedure. Simulations are given in Section 4, and a real-data example is provided in Section 5. Concluding remarks are offered in Section 6. Proofs are relegated to the Appendices, which can be found in the Supplementary Material.

The following notation is used throughout the paper. For a matrix A of size $n \times n$, A_{ij} is its (i, j) th element, $\text{tr}(A)$ is its trace, A' is its transpose, $\det(A)$ is its determinant, $\rho(A)$ is its spectral radius, $\|A\| = \text{tr}(A'A)^{1/2}$ is its Frobenius

norm, $\text{vec}(A)$ is a vector obtained by stacking all the columns of A , $\text{vech}(A)$ is a vector obtained by stacking all columns of the lower triangular part of A , D_n is the $n^2 \times n(n+1)/2$ duplication matrix defined by $D_n \text{vech}(A) = \text{vec}(A)$, $D_n^+ = (D_n' D_n)^{-1} D_n'$ is its generalized inverse, L_n is the elimination matrix defined by $L_n \text{vec}(A) = \text{vech}(A)$, and K_{nn} is the transposition-permutation matrix defined by $K_{nn} \text{vec}(A) = \text{vec}(A')$. For a random variable ξ , $\|\xi\|_k = (E|\xi|^k)^{1/k}$ is its L^k -norm. In addition, I_n is the identity matrix of size $n \times n$, $\mathbf{1}(\cdot)$ is the indicator function, C is a universal constant, $A \otimes B$ is the Kronecker product of two matrices A and B , $A^{\otimes 2} = A \otimes A$, $\rightarrow_{\mathcal{L}}$ denotes convergence in distribution, and the range of integration is $[-1, 1]$, unless stated otherwise.

2. Testing for Structural Changes

2.1. The testing model and hypotheses

Let $\mathbf{y}_t \in \mathbb{R}^{n \times n}$ for $t = 1, 2, \dots, T$ be a positive-definite matrix-valued time series. We consider the following data-generating process (DGP) for \mathbf{y}_t :

$$\mathbf{y}_t = \Sigma_t^{1/2} \mathbf{u}_t \Sigma_t^{1/2} \text{ with } \Sigma_t = \Sigma \left(\frac{t}{T} \right), \quad (2.1)$$

where $\Sigma(x)$ is a deterministic bounded positive-definite matrix function with unknown form on the interval $[0, 1]$, and $\{\mathbf{u}_t\}_{t=1}^T$ is a sequence of strictly stationary positive definite random matrices with $E\mathbf{u}_t = I_n$. In model (2.1), both the dynamic process \mathbf{u}_t and its parameter matrix Σ_t have unspecified forms, allowing the structure of \mathbf{y}_t to change abruptly or smoothly. Our idea of using a time-varying function Σ_t to capture the structural change is motivated by the pioneering work of Robinson (1989, 1991) for the linear regression model with time-varying parameters. Here, we adopt a multiplicative form in model (2.1) to maintain the positive-definite feature of \mathbf{y}_t . Moreover, note that the possible nonstationarity of \mathbf{y}_t in model (2.1) is caused by the deterministic Σ_t . This setting is in accordance with the semiparametric model in Section 3, but does not allow Σ_t to be nonstationary driven by some unit root-type process. To study the unit root-type nonstationarity of \mathbf{y}_t , one may refer to the testing and modelling methodologies and their related technical treatments in Wang and Phillips (2012) and Chan and Wang (2015).

In the literature, researchers tend to build parametric models for the RCOV matrix by implicitly assuming the stationarity of \mathbf{y}_t , or equivalently, the constancy of Σ_t . However, changes are common, and tend to be the rule, rather than the exception. It is well documented that ignoring structural changes can

mislead conventional time series analysis procedures, resulting in erroneous conclusions (Pesaran and Timmermann (2004)). This motivates us to propose a formal structural change test for model (2.1). Our other motivation for the structural change test is triggered by the spurious long-memory phenomenon (see the Supplementary Material for numerical evidence). However, this important aspect has been overlooked in matrix time series applications. For instance, RCOV matrix time series data often exhibit the long-memory phenomenon, and empirical researchers usually apply long-memory models to fit RCOV matrix data (Corsi (2009); Chiriac and Voev (2011)). We argue that these fitted long-memory models should be used with caution, because they do not consider the possibility of spurious long-memory phenomena. Motivated by these arguments, the null hypothesis of interest is

$$H_0 : \Sigma_t = \Sigma_0 \text{ for all } t,$$

where Σ_0 is a positive-definite constant matrix in $\mathbb{R}^{n \times n}$. The alternative hypothesis H_1 is that H_0 is false. Under H_0 , the structure of \mathbf{y}_t is unchanged over time. Under H_1 , Σ_t is a time-varying parameter matrix, and allows the structure of \mathbf{y}_t to change smoothly or abruptly. Testing for structural changes has been well studied for regression or conditional mean time series models; see Fu and Hong (2019) and the references therein. For conditional variance models, most efforts focus on GARCH models; see Berkes, Horváth and Kokoszka (2004) and Fryzlewicz and Rao (2014) for abrupt structural change alternatives, Amado and Teräsvirta (2008) for smooth structural change alternatives, and Chen and Hong (2016) for both abrupt and smooth structural change alternatives. For multivariate conditional variance models with an unknown form, related structural change tests are absent. This is also true for our matrix model (2.1) with $n > 1$.

2.2. Test statistic

We adopt the idea of the generalized Hausman test in Chen and Hong (2012) to propose a structural change test for model (2.1). Here, we need two estimators of Σ_t , where one estimator is always consistent and the other is consistent only under H_0 . Then, our new test measures the distance of these two estimators under a certain norm to look for evidence of rejection.

Our first estimator of Σ_t is chosen as

$$\bar{\Sigma} = \frac{1}{T} \sum_{s=1}^T \mathbf{y}_s,$$

which is obviously efficient and consistent under H_0 , but inconsistent under H_1 .

To introduce our second estimator of Σ_t , we simply consider the Nadaraya–Watson estimator of $\Sigma(x)$, $x \in [0, 1]$, given by

$$\check{\Sigma}(x) = \frac{\sum_{s=1}^T K_h(x - s/T) \mathbf{y}_s}{\sum_{s=1}^T K_h(x - s/T)}$$

(see, e.g., Fan and Yao (2008) for other choices of kernel estimator), where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function, and h is the bandwidth. Because $(1/T) \sum_{s=1}^T K_h(x - s/T) = 1 + O(T^{-1}h^{-1})$ by the integrability of $K(\cdot)$, it is more convenient to consider an alternative estimator of $\Sigma(x)$ given by

$$\hat{\Sigma}(x) = \frac{1}{T} \sum_{s=1}^T K_h\left(x - \frac{s}{T}\right) \mathbf{y}_s. \quad (2.2)$$

Given that $E\mathbf{y}_t = \Sigma_t$ in model (2.1), we then take our second estimator of Σ_t as $\hat{\Sigma}_t = \hat{\Sigma}(t/T)$.

A practical issue for $\hat{\Sigma}_t$ is the edge effect in finite samples, that is, there is insufficient information for the estimators at the extreme left and right ends of the sample interval. To avoid such a problem, following Chen and Hong (2012), we consider the reflection method proposed by Hall and Wehrly (1991) to generate pseudo data $\mathbf{y}_t = \mathbf{y}_{-t}$ for $-[Th] \leq t \leq -1$, and $\mathbf{y}_t = \mathbf{y}_{2T-t}$ for $T + 1 \leq t \leq T + [Th]$. Using the reflection method, we make the boundary points act similarly to those in the interior region $[Th] < t < T - [Th]$, and modify $\hat{\Sigma}_t$ as

$$\hat{\Sigma}(x) = \frac{1}{T} \sum_{s=t-[Th]}^{t+[Th]} K_h\left(x - \frac{s}{T}\right) \mathbf{y}_s, \quad (2.3)$$

for each $t \in \{1, 2, \dots, T\}$.

Another way to circumvent the edge effect is to use the local linear estimator (Fan and Gijbels (1996)), for which the intercepts and slopes are estimated via the weighted least squares. We feel that the arguments used in Fu and Hong (2019) still apply to the test statistic defined in (2.4), with some appropriate modifications. However, using the local linear estimator complicates the estimation theory for the proposed semiparametric model in Section 3. Hence, we focus on the kernel (local constant) estimator, for brevity, and leave the investigation of the local linear estimator for future research.

With the two estimators $\bar{\Sigma}$ and $\hat{\Sigma}_t$, we measure their distance in the Frobenius norm (cumulatively over all t) by

$$\hat{S} = \frac{1}{T} \sum_{t=1}^T [\text{vech}(\hat{\Sigma}_t - \bar{\Sigma})' \text{vech}(\hat{\Sigma}_t - \bar{\Sigma})].$$

Formally, our test statistic is a standardized version of \hat{S} , defined by

$$\hat{D} = \frac{Th^{1/2}\hat{S} - \hat{B}}{\hat{V}^{1/2}}, \tag{2.4}$$

where $\hat{B} = h^{-1/2}\text{tr}(\hat{M})[\int K^2(x)dx]$, $\hat{V} = 2\text{tr}(\hat{M}^2) \int [\int K(x)K(x + \lambda)dx]^2 d\lambda$, and

$$\hat{M} = \sum_{j=-b_T}^{b_T} k\left(\frac{j}{b_T}\right)\hat{\Gamma}_{v,j}.$$

Here, $k(\cdot)$ is another kernel function, b_T is the truncated lag, and

$$\hat{\Gamma}_{v,j} = \begin{cases} \frac{1}{T} \sum_{t=j+1}^T \hat{v}_t \hat{v}'_{t-j}, & \text{for } j = 0, 1, \dots, b_T, \\ \frac{1}{T} \sum_{t=1-j}^T \hat{v}_{t+j} \hat{v}'_t, & \text{for } j = -1, -2, \dots, -b_T, \end{cases}$$

with $\hat{v}_t = \text{vech}(\mathbf{y}_t - \bar{\Sigma})$.

In (2.4), the factors \hat{B} and \hat{V} are the estimators of the mean and variance, respectively, of $Th^{1/2}\hat{S}$. Both of them depend on \hat{M} , which is essentially the estimator for the asymptotic variance of $T^{-1/2} \sum_{t=1}^T \text{vech}(\mathbf{y}_t - \Sigma_0)$ (see Newey and West (1987); Andrews (1991)). Note that the test in Chen and Hong (2012) does not need to consider the corresponding long-run variance estimator, because their assumed model error is a martingale difference sequence (m.d.s.). Fu and Hong (2019) later relaxed this m.d.s. assumption with the help of instrumental variables. In our model (2.1), we only require $E(\mathbf{u}_t - I_n) = 0$ instead of the m.d.s. assumption that $E(\mathbf{u}_t - I_n | \mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_t = \sigma(\mathbf{y}_s, s \leq t)$ is the sigma field generated by the information up to time t . Hence, our test statistic \hat{D} is robust to not only conditional heteroscedasticity and higher-order moments of unknown form, as in Chen and Hong (2012), but also endogeneity, as in Fu and Hong (2019).

2.3. Asymptotic theory

To study the limiting distribution of \hat{D} , the following four assumptions are needed.

Assumption 1. $\Sigma(x)$ is bounded, positive-definite, and continuous, except at a finite number of points on $[0, 1]$, such that for some constants $0 < c_l \leq c_u < \infty$, $c_l \leq \inf_{x \in [0, 1]} \|\Sigma(x)\| \leq \sup_{x \in [0, 1]} \|\Sigma(x)\| \leq c_u$.

Assumption 2. (i) $\text{vech}(\mathbf{u}_t)$ is a strictly stationary β -mixing process with mixing coefficients $\beta(j)$ satisfying $\sum_{j=1}^{\infty} j^2 \beta(j)^{\delta/(1+\delta)} < \infty$, for some $0 < \delta < 1$; (ii) $E \|\mathbf{u}_t\|^{4(1+\delta)} < \infty$.

Assumption 3. (i) $K : [-1, 1] \rightarrow \mathbb{R}^+$ is symmetric about zero, bounded, and twice continuously differentiable, with $\int K(x) dx = 1$ and $\int x^r K(x) dx = C_r < \infty$, for $r \geq 2$; (ii) $h = c_h T^{-\lambda_h}$, for some $0 < \lambda_h < 1/2$ and $0 < c_h < \infty$.

Assumption 4. (i) $k : [-1, 1] \rightarrow \mathbb{R}^+$ is symmetric about zero, bounded, and square integrable, with $k(0) = 1$; (ii) $b_T = c_b T^{\lambda_b}$, where $0 < c_b < \infty$ and $\lambda_b/2 < \lambda_b < 1/2 - 1/2\lambda_h$, for some $0 < \lambda_h < 1/2$.

Assumption 1 is sufficient to guarantee that $\Sigma(x)$ is integrable on $[0, 1]$ up to any finite order, and when $n = 1$, similar conditions have been used in Cavaliere and Taylor (2007), Xu and Phillips (2008), Zhu (2019), and many others. Assumptions 2–4 are in line with Fu and Hong (2019). Assumption 2(i) requires some sufficient technical conditions on \mathbf{u}_t . The stationarity condition for \mathbf{u}_t might be restrictive under some circumstances. Future research may apply the techniques in Zhang and Wu (2012) and Vogt (2015) to relieve this condition. The β -mixing condition is widely adopted in the literature. As shown in Proposition 1 below, Assumption 2(i) holds when \mathbf{u}_t follows model (3.1), with $\rho(\mathcal{A}_0 + \mathcal{B}_0) < 1$. Assumption 2(ii) imposes a higher moment condition on \mathbf{u}_t so that the limit of \widehat{M} under H_0 (denoted by M) exists and some mixing inequalities can be applied, where

$$M = \Sigma_{0*} Z_{\infty} \Sigma'_{0*} \quad (2.5)$$

with $\Sigma_{0*} = (\Sigma_0^{1/2})^{\otimes 2} D_n$, and

$$Z_{\infty} = \sum_{j=-\infty}^{\infty} E(z_t z'_{t-j}) \quad (2.6)$$

with $z_t = \text{vech}(\mathbf{u}_t - I_n)$. Assumption 3(i) holds for commonly used kernels, such as the uniform, Epanechnikov, and truncated Gaussian kernels, among many others. Assumption 3(ii) requires that h converges to zero at a slower rate than $T^{-1/2}$. Assumption 4(i) allows for Bartlett, Parzen, and Tukey–Hanning kernels or the truncated quadratic-spectral (QS) kernel, where the truncated QS kernel is defined by restricting the domain of the QS kernel to $[-1, 1]$. Assumption 4(ii)

gives a sufficient technical condition on b_T to ensure the consistency of \widehat{M} under the null, and was adopted in Hong, Wang and Wang (2017) and Fu and Hong (2019).

Now, we are ready to give the limiting null distribution of \widehat{D} .

Theorem 1. *Suppose Assumptions 1–4 hold. Then, under H_0 ,*

$$\widehat{D} \rightarrow_{\mathcal{L}} N(0, 1) \text{ as } T \rightarrow \infty.$$

Let c_α be the α -upper percentile of $N(0, 1)$. Because a large value of \widehat{D} indicates a rejection of H_0 , we set the critical region of our test as

$$\widehat{D} > c_\alpha, \tag{2.7}$$

which has an asymptotic significance level α , by Theorem 1.

To compute \widehat{D} , we need to select two kernel functions K and k , the bandwidth h , and the truncated lag b_T . As demonstrated in our numerical studies, the choice of K seems less sensitive, and a simple rule-of-thumb of $h = T^{-1/5}/\sqrt{12}$, as in Fu and Hong (2019), works well. Here, the constant $1/\sqrt{12}$ is the standard deviation of a uniform distribution on $[0, 1]$, which can be viewed as the limiting distribution of the grid points t/T , for $t = 1, 2, \dots, T$. Typically, one would expect that the constant c_h in $h = c_h T^{-1/5}$ should depend on the structure of the underlying process \mathbf{y}_t and the smoothness of $\Sigma(x)$. In the framework of estimation, c_h can be optimally chosen by minimizing the integrated mean square error (Robinson (1989)). However, the chosen c_h may not be optimal for the test, as pointed out by Hong, Wang and Wang (2017). In general, to obtain the optimal bandwidth for \widehat{D} , we may need higher-order Edgeworth expansions for the finite-sample distribution of \widehat{D} , and we leave this interesting topic for future work.

To select the kernel function k and the truncated lag b_T , we use the method in Andrews (1991) so that the mean and variance of $Th^{1/2}\widehat{S}$ can be estimated precisely. Specifically, we consider the truncated QS kernel for k with $b_T = 1.3221(\pi T)^{1/5}$, where π is a function of the unknown spectral density matrix of $\text{vech}(\mathbf{y}_t - \Sigma_0)$. We use the truncated QS kernel, because its related long-run variance estimator is close to the one based on the QS kernel for large b_T , and Andrews (1991) has shown that the QS kernel with the preceding choice of b_T is optimal under the asymptotic truncated mean squared error optimality criterion. In practice, one can estimate π by $\widehat{\pi}$ using preliminary information. For instance, we can take $\widehat{\pi} = \min(\widehat{\pi}_1, \widehat{\pi}_2)$ with the upper bound $\widehat{\pi}_1 = 1600$, as suggested by

Lima and Xiao (2010), and

$$\hat{\pi}_2 = \frac{\sum_{a=1}^{n(n+1)/2} 4(1 + \hat{\phi}_a \hat{\psi}_a)^2 (\hat{\phi}_a + \hat{\psi}_a)^2 \hat{\sigma}_a^4}{(1 - \hat{\phi}_a)^6 (1 + \hat{\phi}_a^2)} \bigg/ \frac{\sum_{a=1}^{n(n+1)/2} (1 + \hat{\psi}_a)^4 \hat{\sigma}_a^4}{(1 - \hat{\phi}_a)^4} \tag{2.8}$$

(see (6.5) in Andrews (1991)), where $(\hat{\phi}_a, \hat{\psi}_a)$ is the least squares estimator of (ϕ, ψ) in the ARMA(1, 1) model $\bar{y}_{a,t} = \phi \bar{y}_{a,t-1} + \psi \tau_{a,t-1} + \tau_{a,t}$, $\hat{\sigma}_a^2$ is the estimator of $\text{Var}(\tau_{a,t})$, and $\bar{y}_{a,t}$ is a th element of $\text{vech}(\mathbf{y}_t - \bar{\Sigma})$. Our numerical studies below show that this choice of $\hat{\pi}$ performs well in finite samples.

Next, we study the power behavior of our test \hat{D} under the local alternative

$$H_{1T} : \Sigma_t = \Sigma_0 + \frac{\Sigma_{1t}}{c_T} \text{ for all } t,$$

where $\Sigma_{1t} = \Sigma_1(t/T)$, the deterministic matrix function $\Sigma_1(x)$ satisfies the conditions for $\Sigma(x)$ in Assumption 1, and

$$\mathcal{B}_l = \int_0^1 \text{vech}(\Sigma_1(x))' \text{vech}(\Sigma_1(x)) dx - \int_0^1 \text{vech}(\Sigma_1(x))' dx \int_0^1 \text{vech}(\Sigma_1(x)) dx > 0.$$

Note that \hat{S} using the Frobenius norm can be viewed as an L_2 -type statistic, and its convergence rate in Theorem 1 is $Th^{1/2}$. This implies that the local rate c_T should have the order $O((Th^{1/2})^{1/2}) = O(T^{1/2}h^{1/4})$ to ensure the nontrivial local power of \hat{D} .

Theorem 2. *Suppose Assumptions 1–3, 4(i) hold, and Assumption 4(ii) holds with $\lambda_h/2 < \lambda_b < 1/2 - \lambda_h$, for some $0 < \lambda_h < 1/3$. Under H_{1T} , as $T \rightarrow \infty$,*

- (i) *if $T^{1/2}h^{1/4}/c_T \rightarrow 0$, then $\hat{D} \rightarrow_{\mathcal{L}} N(0, 1)$;*
- (ii) *if $T^{1/2}h^{1/4}/c_T \rightarrow c \in (0, \infty)$, then $\hat{D} \rightarrow_{\mathcal{L}} N(c^2 \mathcal{B}_{lv}, 1)$;*
- (iii) *if $T^{1/2}h^{1/4}/c_T \rightarrow \infty$, then $\hat{D} \rightarrow \infty$ in probability;*

where $\mathcal{B}_{lv} = \mathcal{B}_l/\sqrt{\mathcal{V}} > 0$, with $\mathcal{V} = 2\text{tr}(M^2) \int [\int K(x)K(x + \lambda)dx]^2 d\lambda$.

The above theorem implies that \hat{D} has nontrivial power against local alternatives with decay rate $T^{1/2}h^{1/4}$; it has no power if c_T is faster than $T^{1/2}h^{1/4}$, and it is consistent if c_T is slower than $T^{1/2}h^{1/4}$. The local rate $T^{1/2}h^{1/4}$ is only slightly slower than the parametric local rate $T^{1/2}$ as $h \rightarrow 0$. For example, if $h \propto T^{-1/5}$, then $T^{1/2}h^{1/4} \propto T^{9/20}$ is close to $T^{1/2}$. The same local rate $T^{1/2}h^{1/4}$ is also considered in Härdle and Mammen (1993), Chen and Hong (2012), Fu and Hong (2019), and many others.

Moreover, the modification of Assumption 4(ii) is tailored to give a sufficient technical condition on b_T to ensure the consistency of \widehat{M} for Theorem 2(i)–(ii) (see Proposition A.5 in the Supplementary Material). Here, the commonly used bandwidth $h \propto T^{-1/5}$ is still allowed under the conditions in Theorem 2.

3. A Semiparametric Matrix Model

3.1. Model specification

Once the null hypothesis H_0 is rejected by the test statistic \widehat{D} , a semiparametric matrix model with an appropriate specification for \mathbf{u}_t is needed to model \mathbf{y}_t . Here, we model \mathbf{u}_t by

$$\begin{aligned} \mathbf{u}_t &= G_t^{1/2} \mathbf{e}_t G_t^{1/2} \\ \text{with } G_t &= I_n - A_0 A_0' - B_0 B_0' + A_0 \mathbf{u}_{t-1} A_0' + B_0 G_{t-1} B_0', \end{aligned} \quad (3.1)$$

where $G_t \in \mathcal{F}_{t-1}$ has the first-order BEKK specification in Engle and Kroner (1995) with $E G_t = I_n$, and $\{\mathbf{e}_t\}_{t=1}^T$ is a sequence of independent and identically distributed (i.i.d.) positive-definite random innovation matrices with $E \mathbf{e}_t = I_n$. With minor modifications, our work can be easily extended to the case where G_t has a higher-order mean-targeted BEKK specification, which is not considered below for ease of presentation.

From now on, we call model (2.1) with $\mathbf{u}_t \sim (3.1)$ the time-varying BEKK (TVB) model. Clearly, as a special case, the TVB model becomes stationary when $\Sigma_t \equiv \Sigma_0$, for all t . In the TVB model, \mathbf{y}_t has a nonparametric long-run component Σ_t and a parametric short-run component G_t , which jointly specify the dynamic structure of \mathbf{y}_t with $E \mathbf{y}_t = \Sigma_t$ and $E(\mathbf{y}_t | \mathcal{F}_{t-1}) = \Sigma_t^{1/2} G_t \Sigma_t^{1/2}$. The idea of using a combination of nonparametric and parametric components was first proposed by Feng (2004) for the univariate volatility model. The idea has since become popular for describing the dynamics of univariate volatility (see, e.g., Engle and Rangel (2008); Xu and Phillips (2008); Chen and Hong (2012); Zhu (2019); Zhang et al. (2020); Jiang, Li and Zhu (2021)) and multivariate volatility (see, e.g., Hafner and Linton (2010); Patilea and Raïssi (2010); Amado and Teräsvirta (2014)). The TVB model can thus be viewed as a matrix generalization of the model in Feng (2004). Furthermore, under some suitable conditions, the TVB model could be approximated by a family of locally stationary processes (Dahlhaus (1997)), meaning that \mathbf{y}_t is close to a stationary process $\widehat{\mathbf{y}}_t(s)$, for t/T in a small neighborhood of s .

3.2. Probabilistic properties

In this subsection, we first study the stationarity of model (3.1). Let $\mathcal{A}_0 = L_n A_0^{\otimes 2} D_n$ and $\mathcal{B}_0 = L_n B_0^{\otimes 2} D_n$. By Theorem 2.1 in Zhou et al. (2022), we have the following proposition, which gives the stationarity and moment properties of \mathbf{u}_t .

Proposition 1. *Suppose (i) \mathbf{e}_t admits a density F_e , which is absolutely continuous w.r.t. the Lebesgue measure; (ii) the point I_n is in the interior of the support of F_e ; (iii) $E \|\mathbf{e}_t\| < \infty$. Then, \mathbf{u}_t is strictly stationary and geometrically ergodic with $E \|\mathbf{u}_t\| < \infty$ if and only if $\rho(\mathcal{A}_0 + \mathcal{B}_0) < 1$.*

Note that if \mathbf{u}_t is geometrically ergodic, it is also β -mixing with exponential decay. Hence, if $\rho(\mathcal{A}_0 + \mathcal{B}_0) < 1$, \mathbf{y}_t is β -mixing with exponential decay by Proposition 1 and the boundedness of $\Sigma(x)$. If we do not require $E \|\mathbf{u}_t\| < \infty$, we may use a milder condition than $\rho(\mathcal{A}_0 + \mathcal{B}_0) < 1$ to ensure that \mathbf{y}_t is β -mixing with exponential decay; see, for example, Hafner and Preminger (2009).

Next, we study the moments of \mathbf{y}_t when $\mathbf{u}_t \sim (3.1)$. Let $u_t = \text{vech}(\mathbf{u}_t)$. By rearranging the terms, u_t admits a vector moving average (VMA) structure: $u_t = \text{vech}(I_n) + \sum_{i=0}^{\infty} \Phi_i r_{t-i}$, where $\Phi_0 = I_n$, $\Phi_i = (\mathcal{A}_0 + \mathcal{B}_0)^{i-1} \mathcal{A}_0$ and $r_t = E(u_t | \mathcal{F}_{t-1}) - u_t$ is an m.d.s. By using this VMA representation, we can get the cross-covariance of \mathbf{u}_t and \mathbf{y}_t ; the details are straightforward, and are hence omitted.

Proposition 2. *Suppose \mathbf{u}_t is strictly stationary, with $E \|\mathbf{u}_t\|^2 < \infty$. Then, (i) $E(z_t z'_{t-j}) = \sum_{i=0}^{\infty} \Phi_{j+i} E(r_t r'_t) \Phi'_i$; (ii) $E(w_t w'_{t-j}) = L_n (\Sigma_t^{1/2})^{\otimes 2} D_n E(z_t z'_{t-j}) D'_n (\Sigma_{t-j}^{1/2})^{\otimes 2} L'_n$, where z_t is defined as in (2.6), and $w_t = \text{vech}(\mathbf{y}_t - \Sigma_t)$.*

3.3. Model estimation

Let $\phi = (\text{vec}(A)', \text{vec}(B)')' \in \Gamma_\phi$ be the unknown parameter in model (3.1), and $\phi_0 = (\text{vec}(A_0)', \text{vec}(B_0)')' \in \Gamma_\phi$ be its true value, where Γ_ϕ is the parameter space. In this subsection, we use a two-step estimation procedure to estimate the nonparametric function $\Sigma(x)$ and the unknown parameter ϕ .

Our procedure first estimates $\Sigma(x)$ by $\hat{\Sigma}(x)$ in (2.2). To study the asymptotic property of $\hat{\Sigma}(x)$, two technical assumptions are needed.

Assumption 5. \mathbf{u}_t is strictly stationary and ergodic, with $E \|\mathbf{u}_t\|^2 < \infty$.

Assumption 6. $h = c_h T^{-\lambda_h}$, for some $1/5 \leq \lambda_h < 1$ and $0 < c_h < \infty$.

Assumption 5 gives some regularity conditions on \mathbf{u}_t , which imply that \mathbf{u}_t is β -mixing with exponential decay by Proposition 1. Assumption 6 requires that

h converges to zero at a slower rate than T^{-1} , and is weaker than Assumption 3(ii).

Let $\sigma(x) = \text{vech}(\Sigma(x))$, $\hat{\sigma}(x) = \text{vech}(\hat{\Sigma}(x))$, $b(x) = C_2(\partial^2\sigma(x)/\partial x^2)/2$, and

$$V_\sigma(x) = \left[\int K^2(x)dx \right] L_n[\Sigma(x)^{1/2}]^{\otimes 2} D_n Z_\infty D_n' [\Sigma(x)^{1/2}]^{\otimes 2} L_n'.$$

Using this notation, the asymptotics of $\hat{\Sigma}(x)$ are given in the following theorem.

Theorem 3. *Suppose Assumptions 1, 3(i), and 5–6 hold. Then, for any $x \in (0, 1)$,*

(i) $\hat{\sigma}(x) \rightarrow (1/2)[\sigma(x-) + \sigma(x+)]$ in probability as $T \rightarrow \infty$, where $\sigma(x-) = \lim_{\bar{x}\uparrow x} \sigma(\bar{x})$ and $\sigma(x+) = \lim_{\bar{x}\downarrow x} \sigma(\bar{x})$;

(ii) furthermore, if $\Sigma(x)$ is twice continuously differentiable on $(0, 1)$, then

$$\sqrt{Th}(\hat{\sigma}(x) - \sigma(x) - h^2b(x)) \rightarrow_{\mathcal{L}} N(0, V_\sigma(x)) \text{ as } T \rightarrow \infty.$$

Remark 1. At continuous points, Theorem 3 shows that $\hat{\Sigma}(x)$ always converges to $\Sigma(x)$, and its asymptotic normality further holds under some smoothness conditions on $\Sigma(x)$. At discontinuous points, the consistency of $\hat{\Sigma}(x)$ fails, but this has no impact on the asymptotics of the estimator of ϕ ; see Theorem 4 and Remark 3 below.

Based on $\hat{\Sigma}(x)$, we estimate Σ_t by $\hat{\Sigma}_t = \hat{\Sigma}(t/T)$. With this plug-in estimate $\hat{\Sigma}_t$, our procedure next estimates ϕ via a profiled quasi maximum likelihood estimator (QMLE). Specifically, we write the parametric G_t as

$$G_t(\phi) = I_n - AA' - BB' + \mathbf{A}u_{t-1}A' + BG_{t-1}(\phi)B', \tag{3.2}$$

for $t = 1, \dots, T$. Let $\gamma = (\phi', \nu)'$. By assuming that $\mathbf{e}_t \sim \nu^{-1}\text{Wishart}(\nu, I_n)$ with degrees of freedom $\nu > n$, the negative log-likelihood function (ignoring constants) of $\{\mathbf{y}_t\}_{t=1}^T$ is

$$L_T(\gamma) = \nu \sum_{t=1}^T \ell(\mathbf{y}_t, \Omega_t(\phi)) + \sum_{t=1}^T c(\mathbf{y}_t, \nu) \triangleq \nu \mathcal{L}_T(\phi) + \sum_{t=1}^T c(\mathbf{y}_t, \nu), \tag{3.3}$$

where $\Omega_t(\phi) = \Sigma_t^{1/2} G_t(\phi) \Sigma_t^{1/2}$, $\ell(\mathbf{y}, \Omega) = \text{tr}(\Omega^{-1}\mathbf{y}) + \log \det(\Omega)$, and

$$c(\mathbf{y}, \nu) = -(\nu - n - 1) \log \det(\mathbf{y}) + \nu n \log(2) + 2 \sum_{i=1}^n \log \Gamma\left(\frac{\nu + 1 - i}{2}\right) - n\nu \log(\nu).$$

Here, ν can be viewed as a nuisance parameter, because our main interest is to estimate ϕ . We consider the Wishart log-likelihood function in the estimation for two reasons. First, the covariance matrix estimator based on Gaussian observations is Wishart distributed. Second, like the Gaussian QMLE for the BEKK volatility model (Engle and Kroner (1995); Jiang, Li and Zhu (2021)), the Wishart profiled QMLE of ϕ_0 does not require any distributional assumption on \mathbf{e}_t for its asymptotics.

Owing to the unobservable Σ_t and \mathbf{u}_t , the calculation of $L_T(\gamma)$ is infeasible. To circumvent this difficulty, we have to replace Σ_t and \mathbf{u}_t by $\widehat{\Sigma}_t$ and $\widehat{\mathbf{u}}_t$, respectively, where $\widehat{\mathbf{u}}_t = \widehat{\Sigma}_t^{-1/2} \mathbf{y}_t \widehat{\Sigma}_t^{-1/2}$. Consequently, we consider the following profiled log-likelihood function:

$$\widehat{L}_T(\gamma) = \nu \sum_{t=1}^T \ell(\mathbf{y}_t, \widehat{\Omega}_t(\phi)) + \sum_{t=1}^T c(\mathbf{y}_t, \nu) \triangleq \nu \widehat{\mathcal{L}}_T(\phi) + \sum_{t=1}^T c(\mathbf{y}_t, \nu), \tag{3.4}$$

where $\widehat{\Omega}_t(\phi) = \widehat{\Sigma}_t^{1/2} \widehat{G}_t(\phi) \widehat{\Sigma}_t^{1/2}$ and

$$\widehat{G}_t(\phi) = I_n - AA' - BB' + A\widehat{\mathbf{u}}_{t-1}A' + B\widehat{G}_{t-1}(\phi)B', \tag{3.5}$$

for $t = 1, \dots, T$, with given initial values $\widehat{G}_0(\phi)$ and $\widehat{\mathbf{u}}_0$.

Let $\widehat{\gamma} = (\widehat{\phi}', \widehat{\nu})'$ be the minimizer of $\widehat{L}_T(\gamma)$, that is,

$$\widehat{\gamma} = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \widehat{L}_T(\gamma), \tag{3.6}$$

where $\Gamma = \Gamma_\phi \times \Gamma_\nu$, and Γ_ν is a compact subset of $\{\nu : \nu > n\}$. Interestingly, we can see that the calculation of $\widehat{\phi}$ is irrelevant to ν , leading to

$$\widehat{\phi} = \underset{\phi \in \Gamma_\phi}{\operatorname{argmin}} \widehat{\mathcal{L}}_T(\phi). \tag{3.7}$$

Although $\widehat{\phi}$ is defined by assuming $\mathbf{e}_t \sim \nu^{-1} \text{Wishart}(\nu, I_n)$, the consistency and asymptotic normality of $\widehat{\phi}$ hold for any distribution of \mathbf{e}_t with $E\mathbf{e}_t = I_n$ (see Theorem 4 below). Hence, we call $\widehat{\phi}$ the profiled QMLE of ϕ_0 .

To study the asymptotics of $\widehat{\phi}$, we need three additional assumptions:

Assumption 7. (i) $\det(A_0) \neq 0$; (ii) $\rho(\mathcal{A} + \mathcal{B}) < 1$; (iii) if $G_t(\phi) = G_t(\phi_0)$ (almost surely) for any $\phi, \phi_0 \in \Gamma_\phi$, then $\phi = \phi_0$.

Assumption 8. Γ_ϕ is a compact subset of \mathbb{R}^{2n^2} , and ϕ_0 is an interior point of Γ_ϕ .

Assumption 9. $h = c_h T^{-\lambda_h}$, for some $1/4 < \lambda_h < 1/2$ and $0 < c_h < \infty$.

Assumption 7(i) rules out the deterministic situation of the BEKK process, Assumption 7(ii) ensures the positive definiteness of $G_t(\phi)$, and Assumption 7(iii) is standard for model identifiability; see Engle and Kroner (1995). Assumption 8 gives some regularity conditions for studying the model estimation. Assumption 9 is the key to deriving the asymptotics of $\hat{\phi}$. A restricted h is required to undersmooth the estimation of Σ_t so that the impact of its bias on the asymptotics of $\hat{\phi}$ can be neglected; see Hafner and Linton (2010), Jiang, Li and Zhu (2021), and Remark 4 below for further discussion.

Let $\xi_t = \text{vec}(\mathbf{e}_t - I_n)$, $\rho_{t,i} = \text{vec}(\partial G_t / \partial \phi_i)' (G_t^{-1/2})^{\otimes 2}$, $\eta_{t,i} = \text{vec}(G_t^{-1}(\partial G_t / \partial \phi_i))'$, $M_j^i = E(\mathbf{u}_{t-j} \otimes G_t^{-1}(\partial G_t / \partial \phi_i) G_t^{-1})$, and

$$F_i = \sum_{j=1}^{\infty} \text{vec}(B_0^{j-1} A_0)' M_j^i [I_n \otimes B_0^{j-1} A_0].$$

Using this notation, define

$$\Upsilon(x) = (F - E\eta_t)[\Sigma(x)^{1/4} \otimes \Sigma(x)^{-1/4}],$$

where $F \in \mathbb{R}^{2n^2 \times n^2}$ is a matrix with i th row F_i , and $\eta_t \in \mathbb{R}^{2n^2 \times n^2}$ is a matrix process with i th row $\eta_{t,i}$; furthermore, define

$$J_{\phi_0} = E(\rho_t \rho_t') \text{ and } Q_{\phi_0} = N + \Psi + H + H', \tag{3.8}$$

where $\rho_t \in \mathbb{R}^{2n^2 \times n^2}$ is a matrix process with i th row $\rho_{t,i}$, and $N \in \mathbb{R}^{2n^2 \times 2n^2}$, $\Psi \in \mathbb{R}^{2n^2 \times 2n^2}$, and $H \in \mathbb{R}^{2n^2 \times 2n^2}$ are defined according to

$$\begin{aligned} \text{vec}(N) &= E(\rho_t^{\otimes 2}) \text{vec}(\text{Var}(\xi_t)), \\ \text{vec}(\Psi) &= \left[\int_0^1 \Upsilon(x)^{\otimes 2} dx \right] \Xi_0^{\otimes 2} E[(G_t^{1/2})^{\otimes 4}] \text{vec}(\text{Var}(\xi_t)), \\ \text{vec}(H) &= \left[\int_0^1 (\Upsilon(x) \otimes I_{2n^2}) dx \right] [\Xi_0 \otimes I_{2n^2}] E[(G_t^{1/2})^{\otimes 2} \otimes \rho_t] \text{vec}(\text{Var}(\xi_t)), \end{aligned}$$

respectively, with $\Xi_0 = (I_{n^2} - A_0^{\otimes 2} - B_0^{\otimes 2})^{-1}(I_{n^2} - B_0^{\otimes 2})$.

Now, we give the consistency and asymptotic normality of $\hat{\phi}$.

Theorem 4. *Suppose Assumptions 1, 3(i), and 5–8 hold, and $\Sigma(x)$ is twice continuously differentiable at continuous points on $(0, 1)$. Then,*

(i) $\hat{\phi} \rightarrow \phi_0$ in probability as $T \rightarrow \infty$;

(ii) furthermore, if $\Sigma(x)$ is continuous everywhere on $(0, 1)$, $E\|\mathbf{u}_t\|^{6(1+2\delta)} < \infty$

for some $\delta > 0$, and Assumption 6 is replaced by Assumption 9,

$$\sqrt{T}(\hat{\phi} - \phi_0) \rightarrow_{\mathcal{L}} N(0, V_{\phi_0}) \quad \text{as } T \rightarrow \infty,$$

where $V_{\phi_0} = J_{\phi_0}^{-1} Q_{\phi_0} J_{\phi_0}^{-1}$, and J_{ϕ_0} and Q_{ϕ_0} are defined in (3.8).

Remark 2. When $n = 1$, $\hat{\phi}$ is adaptive to the unknown form of $\Sigma(x)$, because its asymptotic variance is invariant regardless of the form of $\Sigma(x)$. However, when $n > 1$, $\hat{\phi}$ is no longer adaptive because of the term $\Upsilon(x)$, unless $\Sigma(x)^{1/4} \otimes \Sigma(x)^{-1/4} \equiv I_n$ (e.g., $\Sigma(x) = \tau(x)I_n$ with $\tau(x) > 0$, for $x \in (0, 1)$). See also the similar findings for the BEKK volatility model in Jiang, Li and Zhu (2021). Therefore, when $n > 1$, if there exists cross-sectional dependence between any two elements of \mathbf{y}_t , the estimation of Σ_t should affect the asymptotic distribution of $\hat{\phi}$. Note that when $n = 1$, Σ_t represents only the scale of \mathbf{y}_t , and has no effect on the dynamics of \mathbf{u}_t . This may be the reason for the adaptiveness of $\hat{\phi}$ in this univariate case.

Remark 3. To prove Theorem 4(ii), we need a higher smooth condition on $\Sigma(x)$. It is not clear whether this condition can be relieved to allow $\Sigma(x)$ to have finite discontinuous points. Numerically, the simulation studies in Section 4 show that the asymptotic normality of $\hat{\phi}$ in Theorem 4(ii) still holds when $\Sigma(x)$ has abrupt changes. Theoretically, a new technique is required to modify Propositions B.5–B.7 in the Supplementary Material. This kind of modification seems challenging, and hence is left for future work.

Remark 4. To obtain the result in Theorem 4(ii), we need Assumption 9 for our undersmoothing technique, which makes the bias from the first step nonparametric estimator $\hat{\Sigma}_t$ have a negligible effect on the asymptotics of $\hat{\phi}$. Because of Assumption 9, the popular choice of bandwidth $h = c_h T^{-1/5}$ is excluded. To relax Assumption 9, one may use bias correction and small bandwidth techniques (see, e.g., Newey, Hsieh and Robins (2004), Aradillas-Lopez, Honoré and Powell (2007), Zhou and Wu (2010), and Cattaneo and Jansson (2018)), which could account for the bias effect from $\hat{\Sigma}_t$ on the asymptotics of $\hat{\phi}$. How to derive the asymptotics of $\hat{\phi}$ based on the bias correction techniques is interesting and deserves future investigation.

In practice, we can estimate V_{ϕ_0} consistently by its sample version. Specifically, let $\hat{G}_t = \hat{G}_t(\hat{\phi})$, $\partial \hat{G}_t / \partial \phi_i = \partial \hat{G}_t(\hat{\phi}) / \partial \phi_i$, $\hat{\mathbf{e}}_t = \hat{G}_t^{-1/2} \hat{\mathbf{u}}_t \hat{G}_t^{-1/2}$ and define $\hat{\eta}_t$, $\hat{\rho}_t$, $\hat{\xi}_t$ accordingly. Based on these quantities, we compute the sample versions of J_{ϕ_0} and M_j^i by $\hat{J} = (1/T) \sum_{t=1}^T \hat{\rho}_t \hat{\rho}_t'$ and $\hat{M}_j^i = (1/T) \sum_{t=j+1}^T \hat{\mathbf{u}}_{t-j} \otimes \hat{G}_t^{-1} (\partial \hat{G}_t / \partial \phi_i) \hat{G}_t^{-1}$,

respectively. To compute the sample version of Q_{ϕ_0} , we first let

$$\hat{\Upsilon}\left(\frac{t}{T}\right) = \left(\hat{F} - \frac{1}{T} \sum_{s=1}^T \hat{\eta}_s\right) [\hat{\Sigma}_t^{1/4} \otimes \hat{\Sigma}_t^{-1/4}],$$

where the i th row of \hat{F} is \hat{F}_i given by

$$\hat{F}_i = \sum_{j=1}^{b_T} \text{vec}(\hat{B}^{j-1} \hat{A})' \hat{M}_j^i [I_n \otimes \hat{B}^{j-1} \hat{A}],$$

and b_T is defined in Assumption 4(ii). Based on $\hat{\Upsilon}(t/T)$, we next compute the sample versions of N , Ψ , and H by

$$\begin{aligned} \text{vec}(\hat{N}) &= \frac{1}{T} \sum_{t=1}^T [\hat{\rho}_t^{\otimes 2}] \text{vec}(\widehat{\text{Var}}(\xi_t)), \\ \text{vec}(\hat{\Psi}) &= \left\{ \frac{1}{T} \sum_{t=1}^T \left[\hat{\Upsilon}\left(\frac{t}{T}\right) \otimes \hat{\Upsilon}\left(\frac{t}{T}\right) \right] \right\} [\hat{\Xi} \otimes I_{2n^2}] \\ &\quad \left\{ \frac{1}{T} \sum_{t=1}^T [(\hat{G}_t^{1/2})^{\otimes 2} \otimes \hat{\rho}_t] \right\} \text{vec}(\widehat{\text{Var}}(\xi_t)), \\ \text{vec}(\hat{H}) &= \left\{ \frac{1}{T} \sum_{t=1}^T \left[\hat{\Upsilon}\left(\frac{t}{T}\right) \otimes I_{2n^2} \right] \right\} \hat{\Xi}^{\otimes 2} \frac{1}{T} \sum_{t=1}^T \left\{ [(\hat{G}_t^{1/2})^{\otimes 4}] \right\} \text{vec}(\widehat{\text{Var}}(\xi_t)), \end{aligned}$$

respectively, where $\widehat{\text{Var}}(\xi_t)$ is the sample variance of $\{\hat{\xi}_t\}$. Further, we compute the sample version of Q_{ϕ_0} by \hat{Q} , where $\hat{Q} = \hat{\Psi} + \hat{H} + \hat{H}' + \hat{N}$. Finally, we obtain the sample version of V_{ϕ_0} as $\hat{V} := \hat{J}^{-1} \hat{Q} \hat{J}^{-1}$.

Based on \hat{V} , we can construct a Wald test statistic

$$\hat{W} = T(\Gamma \hat{\phi} - \iota)' (\Gamma \hat{V} \Gamma')^{-1} (\Gamma \hat{\phi} - \iota) \tag{3.9}$$

to test the linear null hypothesis $H_0 : \Gamma \phi_0 = \iota$, where $\Gamma \in \mathbb{R}^{s \times 2n^2}$ is a constant matrix with rank s , and $\iota \in \mathbb{R}^{s \times 1}$ is a constant vector. Conventionally, we reject this linear null hypothesis if \hat{W} is larger than the upper-tailed critical value of χ_s^2 .

It is also of interest to estimate the nuisance parameter ν by $\hat{\nu}$ in (3.6), if we have prior information that $\mathbf{e}_t \sim \nu_0^{-1} \text{Wishart}(\nu_0, I_n)$. Because the calculations of $\hat{\phi}$ and $\hat{\nu}$ in (3.6) are independent, we can first obtain $\hat{\phi}$ via (3.7), and then

compute $\hat{\nu}$ by

$$\hat{\nu} = \underset{\nu \in \Gamma_\nu}{\operatorname{argmin}} \hat{L}_T(\hat{\phi}, \nu). \quad (3.10)$$

The asymptotics of $\hat{\nu}$ are given as follows:

Theorem 5. *Suppose $\mathbf{e}_t \sim \nu_0^{-1} \operatorname{Wishart}(I_n, \nu_0)$ and the conditions in Theorem 4 hold. Then,*

(i) $\hat{\nu} \rightarrow \nu_0$ in probability as $T \rightarrow \infty$;

(ii) $\sqrt{T}(\hat{\nu} - \nu_0) \rightarrow_{\mathcal{L}} N(0, 2J_{\nu_0}^{-1})$ as $T \rightarrow \infty$, where $J_{\nu_0} = (1/2) \sum_{i=1}^n \psi'((\nu_0 + 1 - i)/2) - n/\nu_0$, and $\psi(\cdot)$ is the digamma function.

By some simple calculations, we can see that when $\mathbf{e}_t \sim \nu_0^{-1} \operatorname{Wishart}(I_n, \nu_0)$, the Cramér–Rao lower bound with respect to ν_0 is $2J_{\nu_0}^{-1}$. Interestingly, Theorem 5(ii) shows that $\hat{\nu}$ can attain the Cramér–Rao lower bound regardless of the inclusion of the nonparametric part.

4. Simulation Studies

In this section, we assess the finite-sample performance of the generalized Hausman test and the two-step estimation procedure by means of some simulation experiments.

4.1. Study on the test

In this subsection, we examine the finite-sample performance of the generalized Hausman test \hat{D} . We generate 1,000 replications of sample size $T = 500, 1000, 1500, 3000$ from the following two DGPs:

$$\text{DGP 1 : The TVB model with } A_0 = \begin{pmatrix} 0.7 & 0 \\ 0.1 & 0.6 \end{pmatrix} \text{ and } B_0 = \begin{pmatrix} 0.6 & 0 \\ -0.2 & 0.65 \end{pmatrix},$$

$$\text{DGP 2 : The TVB model with } A_0 = \begin{pmatrix} 0.5 & 0 \\ 0.2 & 0.4 \end{pmatrix} \text{ and } B_0 = \begin{pmatrix} 0.5 & 0.1 \\ -0.2 & 0.7 \end{pmatrix},$$

where the function $\Sigma(x)$ is designed as follows:

$$[\text{No change}] \Sigma(x) = 2I_2, \quad (4.1)$$

$$[\text{Diagonal smooth change}] \Sigma(x) = \left(2 + \frac{rx}{10}\right) I_2, \quad (4.2)$$

$$[\text{Off diagonal smooth change}] \Sigma(x) = 2I_2 + \left(\frac{rx}{10}\right) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (4.3)$$

$$\text{[Single abrupt change]} \Sigma(x) = \left(2 + \frac{r}{10} \mathbf{1} \left(\frac{1}{5} < x \right) \right) I_2, \tag{4.4}$$

$$\text{[Multiple abrupt changes]} \Sigma(x) = \left(2 + \frac{r}{10} \mathbf{1} \left(\frac{1}{5} < x < \frac{4}{5} \right) \right) I_2, \tag{4.5}$$

with $r = 1, \dots, 10$; and the error \mathbf{e}_t satisfies the following two distributions:

$$\text{[Wishart distribution]} \mathbf{e}_t \sim \frac{1}{\nu_0} \text{Wishart}(\nu_0, I_n),$$

$$\text{[matrix-F distribution]} \mathbf{e}_t \sim F \left(\nu_0^\dagger, \frac{\nu_{20} - n - 1}{\nu_{10}} I_n \right) \text{ with } \nu_0^\dagger = (\nu_{10}, \nu_{20}),$$

where we choose $\nu_0 = 20$ for DGP 1, $\nu_0 = 5$ for DGP 2, and $\nu_0^\dagger = (10, 6)$ for both DGPs, and generate $F(\nu_0^\dagger, ((\nu_{20} - n - 1)/\nu_{10})I_n)$ distributed \mathbf{e}_t by

$$\mathbf{e}_t = \left(\frac{\nu_{20} - n - 1}{\nu_{10}} \right) L_t^{1/2} R_t^{-1} L_t^{1/2},$$

with $L_t \sim \text{Wishart}(\nu_{10}, I_n)$ and $R_t \sim \text{Wishart}(\nu_{20}, I_n)$ (see Konno (1991)). Here, the specification of $\Sigma(x)$ in (4.1) is used for the size study, and the specifications of $\Sigma(x)$ in (4.2)–(4.5) are used for the power study.

For each replication, we compute \hat{D} according to the statements in Subsection 2.3. Specifically, we consider three choices of kernel function $K(x)$:

$$\text{[Uniform (U)] } K(x) = \frac{1}{2} \mathbf{1}(|x| \leq 1),$$

$$\text{[Epanechnikov (E)] } K(x) = \frac{3}{4} (1 - x^2) \mathbf{1}(|x| \leq 1),$$

$$\text{[Truncated Normal (TN)] } K(x) = \frac{\exp(-x^2) - \exp(-1/2)}{\text{erf}(1/\sqrt{2})\sqrt{2\pi} - 2\exp(-1/2)} \mathbf{1}(|x| \leq 1),$$

and take the bandwidth $h = (1/\sqrt{12})T^{-1/5}$, where $\text{erf}(\cdot)$ is the error function. We choose the kernel function $k(x)$ to be

$$\text{[Truncated QS (TQS)] } k(x) = \frac{25}{12\pi^2 x^2} \left[\frac{\sin(6\pi/5)}{6\pi x/5} - \cos(6\pi x/5) \right] \mathbf{1}(|x| \leq 1),$$

and the truncated lag $b_T = 1.3221(\hat{\pi}T)^{1/5}$, where $\hat{\pi} = \min(1600, \hat{\pi}_2)$ and $\hat{\pi}_2$ is defined in (2.8). When the value of \hat{D} falls into the critical region in (2.7), we reject H_0 at level α .

Because the size performance of \hat{D} may be a bit sensitive to the choice of h , we follow the suggestion in Fu and Hong (2019) to construct a bootstrapped

critical region, and make decisions using the following moving block bootstrap method:

1. Choose a block length l_T^B such that $l_T^B \rightarrow \infty$ and $l_T^B/T \rightarrow 0$ as $T \rightarrow \infty$;
2. Divide the original data into $T - l_T^B + 1$ blocks $\{\Xi_t\}_{t=1}^{T-l_T^B+1}$, where $\Xi_t = \{\mathbf{y}_t, \dots, \mathbf{y}_{t+l_T^B-1}\}$;
3. Resample $\{\Xi_t\}_{t=1}^{T-l_T^B+1}$ with replacement to form the bootstrapped data $\{\Xi_t^*\}_{t=1}^{T-l_T^B+1}$;
4. Calculate $\widehat{\mathcal{D}}^*$ based on the bootstrapped data in step 3;
5. Repeat steps 3–4 J times to get the bootstrapped test statistics $\{\widehat{\mathcal{D}}_j^*\}_{j=1}^J$, calculate the empirical $100(1 - \alpha)$ sample percentile of $\{\widehat{\mathcal{D}}_j^*\}_{j=1}^J$ (denoted by $\widehat{\mathcal{D}}_\alpha^*$), and set the bootstrapped critical region as

$$\widehat{\mathcal{D}} > \widehat{\mathcal{D}}_\alpha^*; \quad (4.6)$$

6. Reject H_0 at level α when the value of $\widehat{\mathcal{D}}$ falls into the critical region in (4.6).

To implement this moving block method, we follow Politis and White (2004) to take $l_T^B = \lfloor (3/2)c_B^2 T \rfloor^{1/3}$, where $c_B = \text{tr}(\sum_{j=-b_T}^{b_T} k(j/b_T) |j| \widehat{\Gamma}_{v,j}) / \text{tr}(\widehat{M})$, and \widehat{M} and $\widehat{\Gamma}_{v,j}$ are defined as in (2.4). Currently, we are unable to provide the consistency for the above bootstrap method, although our simulation results indicate that it works well. This is because existing techniques for proving bootstrap consistency in parametric models (see, e.g., Künsch (1989)) are not directly transferrable to the case of a nonparametric kernel estimation.

Table 1 reports the sizes of $\widehat{\mathcal{D}}$, where $\widehat{\mathcal{D}}^{AC}$ and $\widehat{\mathcal{D}}^{BC}$ denote the results based on the asymptotic critical region in (2.7) and the bootstrapped critical region in (4.6), respectively. From this table, we find that (i) when the sample size T is small, $\widehat{\mathcal{D}}^{AC}$ is undersized, but the size of $\widehat{\mathcal{D}}^{BC}$ is more accurate; (ii) both $\widehat{\mathcal{D}}^{AC}$ and $\widehat{\mathcal{D}}^{BC}$ have satisfactory size performance, regardless of the choice of $K(x)$ and the distribution of \mathbf{e}_t , when the sample size T increases, and $\widehat{\mathcal{D}}^{BC}$ has slightly better size performance than $\widehat{\mathcal{D}}^{AC}$, especially at level 10%; and (iii) the size of $\widehat{\mathcal{D}}^{BC}$ is slightly stable and close to the nominal level under both the Wishart and the matrix-F distributions.

Next, we examine the power performance of $\widehat{\mathcal{D}}$ under the alternative settings in (4.2)–(4.5). As before, we consider both $\widehat{\mathcal{D}}^{AC}$ and $\widehat{\mathcal{D}}^{BC}$. Their results at level 5% are reported in Fig 1 for the data generated from DGP 1, where $T = 1500$

Table 1. Sizes of \hat{D}^{AC} and \hat{D}^{BC} at level $\alpha = 5\%$ and 10% .

T	$K(x)$	$e_t \sim \text{Wishart}$				$e_t \sim \text{matrix-F}$			
		\hat{D}^{AC}		\hat{D}^{BC}		\hat{D}^{AC}		\hat{D}^{BC}	
		5%	10%	5%	10%	5%	10%	5%	10%
DGP 1									
500	U	0.025	0.045	0.026	0.083	0.038	0.073	0.068	0.115
	E	0.020	0.042	0.045	0.090	0.024	0.038	0.036	0.065
	TN	0.019	0.043	0.044	0.088	0.023	0.038	0.037	0.065
1,000	U	0.040	0.073	0.042	0.092	0.041	0.070	0.051	0.105
	E	0.042	0.060	0.051	0.092	0.022	0.043	0.036	0.077
	TN	0.042	0.059	0.050	0.093	0.022	0.040	0.038	0.079
1,500	U	0.058	0.104	0.049	0.101	0.045	0.080	0.047	0.096
	E	0.043	0.089	0.047	0.099	0.042	0.072	0.051	0.095
	TN	0.041	0.087	0.046	0.101	0.042	0.074	0.052	0.098
3,000	U	0.055	0.101	0.042	0.102	0.049	0.080	0.048	0.097
	E	0.053	0.094	0.045	0.105	0.045	0.079	0.045	0.094
	TN	0.053	0.098	0.044	0.099	0.046	0.079	0.045	0.088
DGP 2									
500	U	0.036	0.067	0.054	0.106	0.018	0.036	0.047	0.080
	E	0.028	0.058	0.052	0.112	0.029	0.059	0.056	0.112
	TN	0.027	0.056	0.053	0.110	0.026	0.061	0.059	0.112
1,000	U	0.052	0.095	0.049	0.103	0.031	0.056	0.036	0.088
	E	0.031	0.065	0.042	0.089	0.025	0.060	0.041	0.089
	TN	0.032	0.066	0.043	0.090	0.024	0.060	0.039	0.086
1,500	U	0.056	0.104	0.053	0.109	0.038	0.076	0.038	0.091
	E	0.046	0.081	0.048	0.100	0.042	0.072	0.042	0.097
	TN	0.047	0.079	0.052	0.105	0.035	0.065	0.043	0.095
3,000	U	0.047	0.083	0.039	0.089	0.043	0.084	0.042	0.087
	E	0.051	0.091	0.047	0.101	0.047	0.076	0.045	0.093
	TN	0.052	0.090	0.048	0.097	0.046	0.077	0.045	0.083

and 3000, $e_t \sim \text{Wishart}$, and \hat{D} is calculated by choosing $K(x) \sim U$. The results for other cases are similar, and hence are not reported. From Fig 1, we find that (i) \hat{D}^{AC} is slightly more powerful than \hat{D}^{BC} ; (ii) the power of \hat{D}^{AC} and \hat{D}^{BC} always increases with r , as expected, and reaches one for large r ; (iii) \hat{D}^{AC} and \hat{D}^{BC} under the off-diagonal smooth (or multiple abrupt) change alternative are, in general, more powerful than the diagonal smooth (or single abrupt) change alternative for any fixed r .

Overall, \hat{D}^{AC} is better than \hat{D}^{BC} in terms of the power to detect both smooth and abrupt changes, while \hat{D}^{BC} shows some advantage over \hat{D}^{AC} in terms of

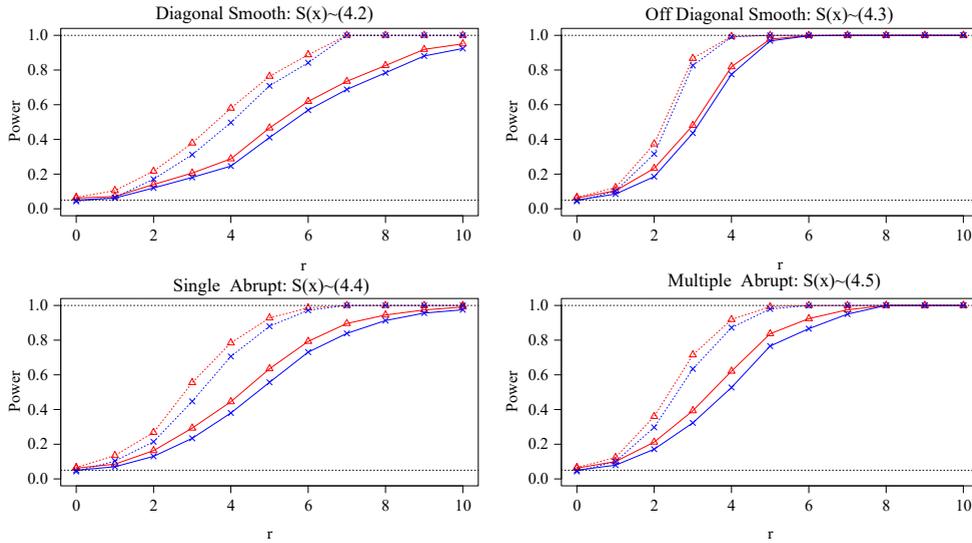


Figure 1. Power across r for \widehat{D}^{AC} (triangle “ Δ ” marker) and \widehat{D}^{BC} (cross “ \times ” marker) when $T = 1500$ (solid line) and $T = 3000$ (dotted line). The horizontal dashed line corresponds to the level $\alpha = 5\%$.

size. Considering the trade-off between size accuracy and power enhancement, we recommend considering the results of both \widehat{D}^{AC} and \widehat{D}^{BC} in practice.

4.2. Study on the estimation

In this subsection, we assess the finite-sample performance of the profiled QMLE $\widehat{\phi}$. We generate 1,000 replications of sample size $T = 1500$ and 3000 from DGPs 1 and 2, where $\Sigma(x)$ is chosen as in (4.2) and (4.5) with $r = 10$. To compute $\widehat{\phi}$ in each repetition, we estimate $\Sigma(x)$ by using $K(x) \sim U$, with a rule-of-thumb bandwidth $h = T^{-1/3}/\sqrt{12}$. Here, we only consider the uniform kernel function for $K(x)$, because the results based on other choices of $K(x)$ are similar. We do not take $h \propto T^{-1/5}$, as suggested by the conventional cross-validation method, because it violates Assumption 9.

Tables 2 and 3 report the sample bias, sample empirical standard deviation (ESD), and average asymptotic standard deviation (ASD) of $\widehat{\phi}$ based on DGPs 1 and 2, respectively, where the ASD is calculated according to Theorem 4(ii). From these two tables, we find that (i) the biases of $\widehat{\phi}$ are small in all considered cases; (ii) regardless of the specification of $\Sigma(x)$ and the distribution of \mathbf{e}_t , the values of ESD and ASD are close to each other (especially for large T); (iii) when the sample size T increases, the values of ESD and ASD decrease as expected; (iv) the values of ESD (or ASD) in the case of $\mathbf{e}_t \sim \text{Wishart}$ are smaller than the

Table 2. Estimation results for $\hat{\phi}$ based on DGP 1.

\mathbf{e}_t	T		A_{11}	A_{21}	A_{12}	A_{22}	B_{11}	B_{21}	B_{12}	B_{22}	ν
Panel A: $\Sigma(x) \sim (4.2)$ with $r = 10$											
Wishart	1,500	Bias	-0.0055	0.0030	-0.0011	-0.0070	-0.0093	-0.0138	0.0102	-0.0090	0.1712
		ESD	0.0172	0.0169	0.0185	0.0189	0.0259	0.0248	0.0285	0.0299	0.4229
		ASD	0.0156	0.0176	0.0180	0.0187	0.0255	0.0244	0.0302	0.0279	0.4089
	3,000	Bias	-0.0021	0.0017	-0.0006	-0.0030	-0.0056	0.0080	0.0051	-0.0047	0.0504
		ESD	0.0118	0.0118	0.0131	0.0133	0.0176	0.0166	0.0203	0.0199	0.2903
		ASD	0.0110	0.0122	0.0126	0.0131	0.0176	0.0166	0.0206	0.0191	0.2886
matrix-F	1,500	Bias	-0.0319	0.0009	0.0019	-0.0245	0.0096	-0.0102	0.0098	-0.0009	\
		ESD	0.0315	0.0306	0.0306	0.0337	0.0436	0.0394	0.0412	0.0413	\
		ASD	0.0296	0.0330	0.0324	0.0409	0.0425	0.0352	0.0384	0.0363	\
	3,000	Bias	-0.0228	0.0012	0.0029	-0.0169	0.0119	-0.0062	0.0028	0.0048	\
		ESD	0.0229	0.0224	0.0224	0.0299	0.0315	0.0269	0.0297	0.0322	\
		ASD	0.0250	0.0236	0.0168	0.0303	0.0309	0.0242	0.0266	0.0248	\
Panel B: $\Sigma(x) \sim (4.5)$ with $r = 10$											
Wishart	1,500	Bias	-0.040	0.0045	-0.0016	-0.0050	-0.0087	-0.0151	0.0124	-0.0670	0.1108
		ESD	0.0172	0.0170	0.0186	0.0190	0.0259	0.0248	0.0287	0.0297	0.4212
		ASD	0.0154	0.0176	0.0179	0.0187	0.0253	0.0240	0.0297	0.0273	0.4085
	3,000	Bias	-0.0011	0.0030	-0.0010	-0.0018	-0.0049	-0.0910	0.0067	0.0198	0.0530
		ESD	0.0118	0.0118	0.0133	0.0133	0.0177	0.0166	0.0205	0.0198	0.2897
		ASD	0.0109	0.0122	0.0125	0.0131	0.0175	0.0164	0.0203	0.0187	0.2880
matrix-F	1,500	Bias	-0.0313	0.0012	0.0019	-0.0239	0.0099	-0.0103	0.0101	-0.0005	\
		ESD	0.0314	0.0306	0.0306	0.0337	0.0436	0.0394	0.0412	0.0472	\
		ASD	0.0195	0.0330	0.0224	0.0410	0.0424	0.0351	0.0382	0.0361	\
	3,000	Bias	-0.0225	0.0014	0.0029	-0.0166	0.0122	-0.0063	0.0030	0.0054	\
		ESD	0.0230	0.0225	0.0224	0.0247	0.0314	0.0269	0.0298	0.0322	\
		ASD	0.0156	0.0236	0.0173	0.0304	0.0308	0.0243	0.0265	0.0252	\

corresponding values of ESD (or ASD) in the case of $\mathbf{e}_t \sim$ matrix-F, and this is consistent with the fact that $\hat{\phi}$ is an efficient estimator when \mathbf{e}_t follows a Wishart distribution. Note that when $\mathbf{e}_t \sim$ Wishart, Tables 2 and 3 also give the results for the estimate $\hat{\nu}$, which performs as well as $\hat{\phi}$.

Overall, our profiled QMLE $\hat{\phi}$ has good finite-sample performance in all considered smooth and abrupt change specifications.

5. An Empirical Example

In this section, we revisit the RCOV matrix time series data on IBM Common Stock (IBM) and Microsoft Corporation (MSFT) in Lunde, Shephard and Sheppard (2016). This data set ranges from January 2006 to December 2011, with 1,474 observations in total. Two flash crashes, on May 6, 2010, and August

Table 3. Estimation results for $\hat{\phi}$ based on DGP 2.

\mathbf{e}_t	T		A_{11}	A_{21}	A_{12}	A_{22}	B_{11}	B_{21}	B_{12}	B_{22}	ν
Panel A: $\Sigma(x) \sim (4.2)$ with $r = 10$											
Wishart	1,500	Bias	-0.0096	0.0055	-0.0012	-0.0111	-0.0492	0.0043	0.0264	-0.0411	0.0437
		ESD	0.0271	0.0290	0.0306	0.0305	0.1003	0.1091	0.1119	0.0887	0.1090
		ASD	0.0263	0.0282	0.0302	0.0293	0.0938	0.0891	0.0983	0.0850	0.1114
	3,000	Bias	-0.0044	0.0037	-0.0017	-0.0045	-0.0249	0.0037	0.0173	-0.0236	0.0225
		ESD	0.0183	0.0201	0.0208	0.0207	0.0623	0.0619	0.0672	0.0581	0.0757
		ASD	0.0184	0.0196	0.0211	0.0205	0.0617	0.0589	0.0653	0.0560	0.0784
matrix-F	1,500	Bias	-0.0177	0.0062	-0.0037	-0.0182	-0.0686	0.0083	0.0460	-0.0509	\
		ESD	0.0421	0.0396	0.0457	0.0438	0.1318	0.1463	0.1334	0.1311	\
		ASD	0.0368	0.0369	0.0386	0.0428	0.1176	0.1195	0.1153	0.1028	\
	3,000	Bias	-0.0091	0.0037	-0.0022	-0.0091	-0.0360	0.0024	0.0233	-0.0252	\
		ESD	0.0312	0.0284	0.0322	0.0319	0.0819	0.0749	0.0783	0.0783	\
		ASD	0.0271	0.0270	0.0289	0.0318	0.0794	0.0732	0.0771	0.0693	\
Panel B: $\Sigma(x) \sim (4.5)$ with $r = 10$											
Wishart	1,500	Bias	-0.0075	0.0064	-0.0028	-0.0089	-0.0370	-0.0079	0.0219	-0.0290	0.0310
		ESD	0.0273	0.0294	0.0312	0.0307	0.0998	0.1097	0.1145	0.0973	0.1090
		ASD	0.0261	0.0279	0.0298	0.0291	0.0886	0.0838	0.0925	0.0797	0.1111
	3,000	Bias	-0.0029	0.0048	-0.0030	-0.0030	-0.0165	-0.0080	0.0141	-0.0129	0.0756
		ESD	0.0184	0.0205	0.0213	0.0208	0.0629	0.0649	0.0689	0.0586	0.0756
		ASD	0.0183	0.0195	0.0209	0.0204	0.0591	0.0560	0.0621	0.0529	0.0783
matrix-F	1,500	Bias	-0.0162	0.0067	-0.0047	-0.0167	-0.0653	0.071	0.0467	-0.0488	\
		ESD	0.0423	0.0398	0.0459	0.0439	0.1325	0.1488	0.1338	0.1343	\
		ASD	0.0367	0.0368	0.0384	0.0428	0.1156	0.1065	0.1130	0.1006	\
	3,000	Bias	-0.0083	0.0051	-0.0028	-0.0085	-0.0349	-0.0063	0.0245	-0.0208	\
		ESD	0.0314	0.0291	0.0323	0.0321	0.0842	0.0853	0.0808	0.0783	\
		ASD	0.0274	0.0271	0.0282	0.0319	0.0781	0.0791	0.0757	0.0674	\

9, 2011, are replaced by the average of their nearest five preceding and following matrices, respectively. Denote this data set by $\{\mathbf{y}_t\}_{t=1}^{1,474}$. First, we plot $\{\mathbf{y}_{rs,t}\}_{t=1}^{1,474}$ and the sample ACFs of $\mathbf{y}_{rs,t}$ in Fig 2, where $\mathbf{y}_{rs,t}$ is the (r, s) th element of \mathbf{y}_t . This figure shows a peak in all $\mathbf{y}_{rs,t}$ around late 2008, and the sample ACFs of each $\mathbf{y}_{rs,t}$ decay slowly.

Next, we apply our generalized Hausman test \hat{D} to detect whether there is any structural change in \mathbf{y}_t . Based on this data set, the value of \hat{D} is 4.2061, where \hat{D} is computed as in Subsection 4.1, with $K(x) \sim U$. Moreover, the p-values of \hat{D} using the asymptotic critical region (as for \hat{D}^{AC}) and the bootstrapped critical region (as for \hat{D}^{BC}) are 0.0000 and 0.0060, respectively. Both methods convey strong evidence that the structure of \mathbf{y}_t is changing over time.

Third, we fit \mathbf{y}_t using our semiparametric TVB model, which is estimated

using the two-step estimation procedure in Subsection 3.3. The top panels of Fig 3 plot the values of $\hat{\Sigma}_{r,s,t}$ (the (r, s) th element of $\hat{\Sigma}_t$), which are computed as in Subsection 4.2. In view of $\hat{\Sigma}_{r,s,t}$, we can see that the long-run component Σ_t is not a constant matrix, and its value from late 2008 to early 2009 is much higher than that of the other time period. This finding is reasonable, because the subprime financial crisis happened in 2008–2009, and the (co-)variance of IBM and MSFT tends to be higher during that period. Using the estimated long-run component, the short-run component G_t in the TVB model is fitted by the profiled QMLE with $\hat{A} = \begin{pmatrix} 0.7227_{(0.0109)} & -0.0119_{(0.0177)} \\ 0.1057_{(0.0187)} & 0.6256_{(0.0279)} \end{pmatrix}$ and $\hat{B} = \begin{pmatrix} 0.5862_{(0.0241)} & 0.0755_{(0.0237)} \\ -0.1962_{(0.0268)} & 0.6799_{(0.0295)} \end{pmatrix}$, where the values in parentheses are the related asymptotic standard errors. Furthermore, we apply the Wald test \hat{W} in (3.9) to examine the two null hypotheses $H'_0 : A_{12,0} = 0$ and $H''_0 : A_{12,0} = B_{12,0} = 0$, and the corresponding p-values are 0.5025 and 0.0013, respectively. Hence, we have strong evidence to reject H''_0 , but not H'_0 . Consequently, we re-fit the short-run component G_t using the profiled QMLE with $\hat{A} = \begin{pmatrix} 0.7212_{(0.0111)} & 0 \\ 0.1004_{(0.0186)} & 0.6274_{(0.0280)} \end{pmatrix}$ and $\hat{B} = \begin{pmatrix} 0.5896_{(0.0243)} & 0.0614_{(0.0234)} \\ -0.1890_{(0.0267)} & 0.6802_{(0.0295)} \end{pmatrix}$. To check the adequacy of our re-fitted TVB model, we plot the first 100 sample ACFs of the residual series $\{\hat{\mathbf{e}}_{r,s,t}\}$ (denoted by $\rho_{rs}(j)$ for $j = 1, \dots, 100$) in the three bottom panels of Fig 3, where $\hat{\mathbf{e}}_{r,s,t}$ is the (r, s) th element of $\hat{\mathbf{e}}_t$. From these three panels, we find that the considered ACFs have small values, in general, except that they exhibit relatively large values at two or three large lags for the cases of $(r, s) = (1, 2)$ and $(2, 2)$. This indicates that our re-fitted TVB model is largely adequate, although the possible dependence within $\mathbf{e}_{12,t}$ and $\mathbf{e}_{22,t}$ could be captured by using high-order or long-memory TVB models.

The conditional heterogeneous autoregressive (HAR) model proposed by Corsi (2009) is used as a benchmark to study the univariate RCOV, because it can capture the observed long-memory feature of the RCOV. As shown in Fig 2, our considered RCOV matrix data also exhibit the long-memory feature, and thus can be fitted by an alternative stationary HAR-type matrix time series model:

$$\begin{aligned} \mathbf{y}_t &= G_t^{1/2} \mathbf{e}_t G_t^{1/2}, \\ G_t &= \Omega + A_{(d)} \mathbf{y}_{t-1,d} A'_{(d)} + A_{(w)} \mathbf{y}_{t-1,w} A'_{(w)} + A_{(m)} \mathbf{y}_{t-1,m} A'_{(m)}, \end{aligned} \quad (5.1)$$

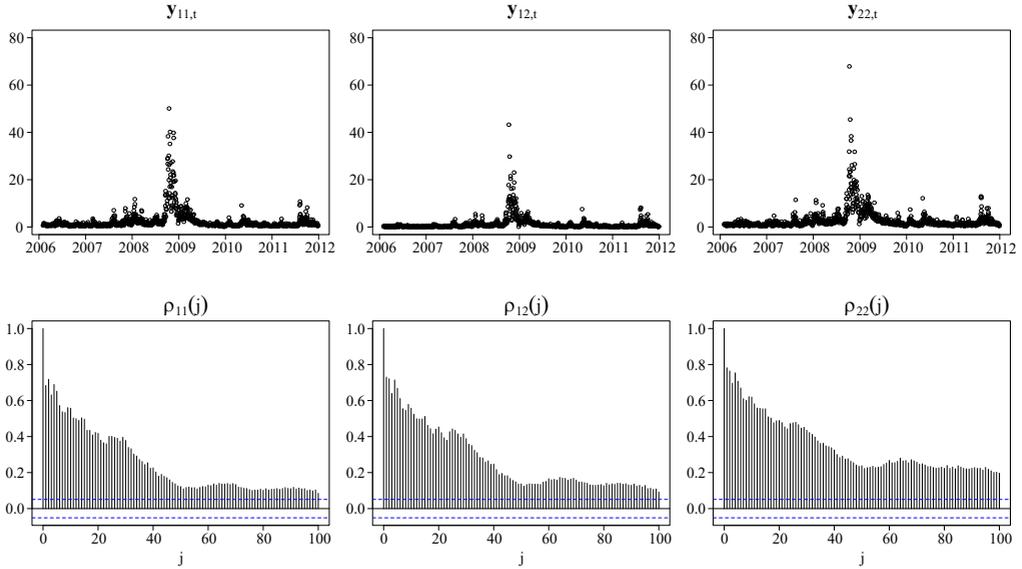


Figure 2. Top panels: $y_{rs,t}$. Bottom panels: sample ACFs of $y_{rs,t}$.

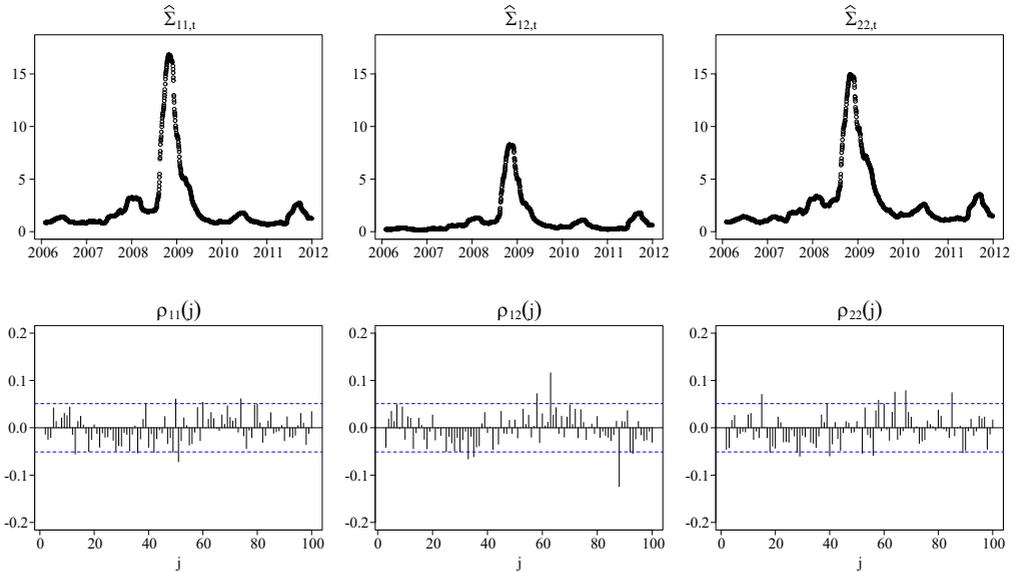


Figure 3. Top panels: $\hat{\Sigma}_{rs,t}$. Bottom panels: sample ACFs of $e_{rs,t}$.

where $\mathbf{y}_{t-1,d} = \mathbf{y}_{t-1}$, $\mathbf{y}_{t-1,w} = (1/5) \sum_{i=1}^5 \mathbf{y}_{t-i}$, and $\mathbf{y}_{t-1,m} = (1/22) \sum_{i=1}^{22} \mathbf{y}_{t-i}$ are the daily, weekly, and monthly averages, respectively, of RCOV matrices. To obtain a fitted model (5.1) for our RCOV matrix data, we estimate the unknown parameters in (5.1) by assuming $\mathbf{e}_t \sim \nu^{-1} \text{Wishart}(\nu, I_n)$. To end this section, we

Table 4. Forecast errors based on TVB and HAR models

	$l = 1$		$l = 5$		$l = 10$	
	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral
TVB model	0.9137	0.8873	1.1978	1.1661	1.3437	1.3087
HAR model	0.9183	0.8900	1.2235	1.1859	1.3808	1.3353

compare the forecasting performance of our TVB model with this HAR model, based on a rolling window procedure with window size equal to 1,000. Specifically, we use the in-sample data $\{\mathbf{y}_t\}_{t=T_0-999}^{T_0}$ to make an l -step-ahead forecast $\hat{\mathbf{y}}_{T_0+l|T_0}$ for the out-of-sample data point \mathbf{y}_{T_0+l} , where $T_0 = 1000, \dots, 1474 - l$. For the TVB model, $\hat{\mathbf{y}}_{T_0+l|T_0}$ is calculated as $\hat{\mathbf{y}}_{T_0+l|T_0} = \hat{\Sigma}_{T_0}^{1/2} \hat{\mathbf{u}}_{T_0+l|T_0} \hat{\Sigma}_{T_0}^{1/2}$, where $\hat{\Sigma}_{T_0}$ is the estimate of Σ_{T_0} , and $\hat{\mathbf{u}}_{T_0+l|T_0}$ is the forecast of \mathbf{u}_{T_0+l} . For the HAR model, $\hat{\mathbf{y}}_{T_0+l|T_0}$ is calculated as $\hat{\mathbf{y}}_{T_0+l|T_0} = \hat{G}_{T_0+l|T_0}$, where $\hat{G}_{T_0+l|T_0}$ is the forecast of G_{T_0+l} . Table 4 reports the average of forecast errors $\{\hat{\mathbf{y}}_{T_0+l|T_0} - \mathbf{y}_{T_0+l}\}$ in the Frobenius and spectral norms for both models, where the forecasting horizon l is taken as 1, 5, and 10, corresponding to daily, weekly, and biweekly forecasts. From this table, we find that regardless of the forecasting horizon, the TVB model always has a smaller forecast error than that of the HAR model. Moreover, applying the DM test in Diebold and Mariano (2002) to our forecasting results, we find that the p-value of the DM test is less than 0.001 in each case, implying that the TVB model has significantly better forecasting accuracy than that of the HAR model. This forecasting advantage of the TVB model over the HAR model may be the result of its ability to take the structural change of \mathbf{y}_t into account.

6. Conclusion

We first provide a generalized Hausman test for detecting the structural change in covariance matrix time series, and then derive its limiting distributions under both null and local alternatives. As illustrated by simulations, matrix time series data can exhibit spurious long-memory phenomena when the structure of their underlying model changes. Our generalized Hausman test is motivated by this, and is applicable without assuming any prior information on the structural change alternative. This makes our test appealing, because there is no guarantee that the structures of the matrix time series specified by the researcher provide a correct description of reality.

We also propose a new semiparametric TVB model that simultaneously allows for structural change and temporal dependence. Our TVB model is estimated by a two-step estimation procedure, and the asymptotics of its related

estimators are established. Because this two-step estimation procedure is valid without specifying the form of the structural change or the distribution of the innovation, it has a wide application scope. By applying our generalized Hausman test and TVB model to one RCOV matrix time series data set, we find strong evidence that this RCOV matrix data set is undergoing a structural change over a BEKK-type model during the examined period. Furthermore, its observed long-memory phenomenon is well captured by the TVB model. Of course, it is possible that the observed long-memory phenomenon in the RCOV data set is caused by other mechanisms, such as the regime-switching mechanism. Hence, we feel that studying the regime-switching matrix time series model is a promising direction for future research. Finally, note that our methodology is for fixed-dimensional covariance matrix time series. Extending our work to the high-dimensional setting (see, e.g., Tao et al. (2011), Leng and Tang (2012), and Wang, Liu and Chen (2019)) could be another promising direction for future research.

Supplementary Material

The Supplementary Material contains Appendices A, B, C, and D, where Appendix A gives proofs of Theorems 1–2, Appendix B offers proofs of Theorems 3–5, Appendix C lists some basic derivatives results, and Appendix D provides some numerical evidence of spurious long-memory phenomena caused by structural changes.

Acknowledgments

We greatly appreciate the helpful comments and suggestions of the anonymous referee, associate editor, and co-editor. We thank Prof. Lijian Yang for pointing out some errors in an earlier version. Jiang acknowledges that some of his work was carried out during visits to the University of Hong Kong. D. Li's work is supported in part by the NSFC (No.11771239 and No.71973077) and Tsinghua University Initiative Scientific Research Program (No.2019Z07L01009). W.K. Li's work is supported in part by Hong Kong GRF grant (No.17304417) and EdUHK grant RG44/2019-2020R. Zhu's work is supported in part by Hong Kong GRF grant (No.17306818, No.17305619, and No.17304421) and the NSFC (No.11690014 and No.11731015).

References

Amado, C. and Teräsvirta, T. (2008). Modelling conditional and unconditional heteroskedasticity with smoothly time-varying structure. Working Paper. *Stockholm School of Economics*.

- Amado, C. and Teräsvirta, T. (2014). Conditional correlation models of autoregressive conditional heteroscedasticity with nonstationary GARCH equations. *Journal of Business & Economic Statistics* **32**, 69–87.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica* **71**, 579–625.
- Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–858.
- Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61**, 821–856.
- Aradillas-Lopez, A., Honoré, B. E. and Powell, J. L. (2007). Pairwise difference estimation with nonparametric control variables. *International Economic Review* **48**, 1119–1158.
- Aue, A., Hörmann, S., Horváth, L. and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics* **37**, 4046–4087.
- Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis* **34**, 1–16.
- Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66**, 47–78.
- Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 253–280.
- Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariance: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* **72**, 885–925.
- Berkes, I., Horváth, L. and Kokoszka, P. (2004). Testing for parameter constancy in GARCH(p, q) models. *Statistics & Probability Letters* **70**, 263–273.
- Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **37**, 149–192.
- Cattaneo, M. D. and Jansson, M. (2018). Kernel-based semiparametric estimators: Small bandwidth asymptotics and bootstrap consistency. *Econometrica* **86**, 955–995.
- Cavaliere, G. and Taylor, A. R. (2007). Testing for unit roots in time series models with nonstationary volatility. *Journal of Econometrics* **140**, 919–947.
- Chan, N. and Wang, Q. (2015). Nonlinear regressions with nonstationary time series. *Journal of Econometrics* **185**, 182–195.
- Chen, B. (2015). Modeling and testing smooth structural changes with endogenous regressors. *Journal of Econometrics* **185**, 196–215.
- Chen, B. and Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica* **80**, 1157–1183.
- Chen, B. and Hong, Y. (2016). Detecting for smooth structural changes in GARCH models. *Econometric Theory* **32**, 740–791.
- Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics* **26**, 922–947.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* **7**, 174–196.
- Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* **25**, 1–37.

- Davis, R. A., Huang, D. and Yao, Y.-C. (1995). Testing for a change in the parameter values and order of an autoregressive model. *The Annals of Statistics* **23**, 282–304.
- Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics* **20**, 134–144.
- Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* **11**, 122–150.
- Engle, R. F. and Rangel, J. G. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies* **21**, 1187–1222.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman & Hall, London.
- Fan, J. and Yao, Q. (2008). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.
- Feng, Y. (2004). Simultaneously modeling conditional heteroskedasticity and scale change. *Econometric Theory* **20**, 563–596.
- Fryzlewicz, P. and Rao, S. S. (2014). Multiple-change-point detection for auto-regressive conditional heteroscedastic processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**, 903–924.
- Fu, Z. and Hong, Y. (2019). A model-free consistent test for structural change in regression possibly with endogeneity. *Journal of Econometrics* **211**, 206–242.
- Golosnoy, V., Gribisch, B. and Liesenfeld, R. (2012). The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics* **167**, 211–223.
- Gouriéroux, C., Jasiak, J. and Sufana, R. (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics* **150**, 167–181.
- Hafner, C. M. and Linton, O. (2010). Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics* **159**, 55–73.
- Hafner, C. M. and Preminger, A. (2009). On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis* **100**, 2044–2054.
- Hall, P. and Wehrly, T. E. (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *Journal of the American Statistical Association* **86**, 665–672.
- Hansen, B. E. (2000). Testing for structural change in conditional models. *Journal of Econometrics* **97**, 93–115.
- Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics* **21**, 1926–1947.
- Hidalgo, J. (1995). A nonparametric conditional moment test for structural stability. *Econometric Theory* **11**, 671–698.
- Hong, Y., Wang, X. and Wang, S. (2017). Testing strict stationarity with applications to macroeconomic time series. *International Economic Review* **58**, 1227–1277.
- Jiang, F., Li, D. and Zhu, K. (2021). Adaptive inference for a semiparametric generalized autoregressive conditional heteroskedasticity model. *Journal of Econometrics* **224**, 306–329.
- Konno, Y. (1991). A note on estimating eigenvalues of scale matrix of the multivariate F -distribution. *Annals of the Institute of Statistical Mathematics* **43**, 157–165.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics* **17**, 1217–1241.

- Leng, C. and Tang, C. Y. (2012). Sparse matrix graphical models. *Journal of the American Statistical Association* **107**, 1187–1200.
- Lima, L. R. and Xiao, Z. (2010). Is there long memory in financial time series? *Applied Financial Economics* **20**, 487–500.
- Lin, C.-F. J. and Teräsvirta, T. (1994). Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics* **62**, 211–228.
- Lunde, A., Shephard, N. and Sheppard, K. (2016). Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice. *Journal of Business & Economic Statistics* **34**, 504–518.
- Mikosch, T. and Stărică, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics* **86**, 378–390.
- Newey, W. K., Hsieh, F. and Robins, J. M. (2004). Twicing kernels and a small bias property of semiparametric estimators. *Econometrica* **72**, 947–962.
- Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703–708.
- Opschoor, A., Janus, P., Lucas, A. and Van Dijk, D. (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics* **36**, 643–657.
- Patilea, V. and Raïssi, H. (2010). Adaptive estimation of vector autoregressive models with time-varying variance: Application to testing linear causality in mean. *Journal of Statistical Planning and Inference* **142**, 2891–2912.
- Pesaran, M. H. and Timmermann, A. (2004). How costly is it to ignore breaks when forecasting the direction of a time series? *International Journal of Forecasting* **20**, 411–425.
- Politis, D. N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* **23**, 53–70.
- Qu, Z. and Perron, P. (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica* **75**, 459–502.
- Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change* (Edited by P. Hackl), 253–264. Springer, Heidelberg.
- Robinson, P. M. (1991). Time-varying nonlinear regression. In *Economic Structural Change*, 179–190. Springer, Heidelberg.
- Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association* **105**, 1228–1240.
- Stărică, C. and Granger, C. (2005). Nonstationarities in stock returns. *Review of Economics and Statistics* **87**, 503–522.
- Su, L. and Xiao, Z. (2008). Testing structural change in time-series nonparametric regression models. *Statistics and its Interface* **1**, 347–366.
- Tao, M., Wang, Y., Yao, Q. and Zou, J. (2011). Large volatility matrix inference via combining low-frequency and high-frequency approaches. *Journal of the American Statistical Association* **106**, 1025–1040.
- Vogt, M. (2015). Testing for structural change in time-varying nonparametric regression models. *Econometric Theory* **31**, 811–859.
- Wang, D., Liu, X. and Chen, R. (2019). Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics* **208**, 231–248.
- Wang, Q. and Phillips, P. C. B. (2012). A specification test for nonlinear nonstationary models. *The Annals of Statistics* **40**, 727–758.

- Xu, K.-L. and Phillips, P. C. B. (2008). Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics* **142**, 265–280.
- Yu, P. L., Li, W. K. and Ng, F. C. (2017). The generalized conditional autoregressive Wishart model for multivariate realized volatility. *Journal of Business & Economic Statistics* **35**, 513–527.
- Zhang, T. and Wu, W. B. (2012). Inference of time-varying regression models. *The Annals of Statistics* **40**, 1376–1402.
- Zhang, Y., Liu, R., Shao, Q. and Yang, L. (2020). Two-step estimation for time varying ARCH models. *Journal of Time Series Analysis* **41**, 551–570.
- Zhou, J., Jiang, F., Zhu, K. and Li, W. K. (2022). Time series models for realized covariance matrices based on the matrix-F distribution. *Statistica Sinica* **32**, 755–786.
- Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**, 513–531.
- Zhu, K. (2019). Statistical inference for autoregressive models under heteroscedasticity of unknown form. *The Annals of Statistics* **47**, 3185–3215.

Feiyu Jiang

Department of Statistics and Data Science, School of Management, Fudan University, Shanghai, 200433, China.

E-mail: jiangfy@fudan.edu.cn

Dong Li

Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing, 100084, China.

E-mail: malidong@tsinghua.edu.cn

Wai Keung Li

Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong.

E-mail: waikeungli@eduhk.hk

Ke Zhu

Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong.

E-mail: mazhuke@hku.hk

(Received January 2021; accepted August 2021)