Statistica Sinica Preprint No: SS-2021-0029	
Title	Testing and Modelling for the Structural Change in Covariance Matrix Time Series With Multiplicative Form
Manuscript ID	SS-2021-0029
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0029
Complete List of Authors	Feiyu Jiang, Dong Li, Wai Keung Li and Ke Zhu
Corresponding Author	Ke Zhu
E-mail	mazhuke@hku.hk
Notice: Accepted version subject to English editing.	

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

BY FEIYU JIANG^{*}, DONG LI[†], WAI KEUNG LI[‡] AND KE ZHU[§]

Fudan University^{}, Tsinghua University[†],*

Education University of Hong Kong[‡] and University of Hong Kong[§]

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 it has non-trivial power in detecting a broad class of alternatives. Moreover, we propose a new semiparametric covariance matrix time series model, which has a time-varying long run component to take the structural change into account, and a BEKK-type short run component to capture the temporal dependence. A two-step estimation procedure is proposed to estimate this semiparametric model, and the asymptotics of the related estimators are established. Finally, the importance of the generalized Hausman test and the semiparametric model is illustrated by simulations and an application to realized covariance matrix data.

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 able to enhance the prediction ability for the covariance matrix of the underlying

Keywords and phrases: Covariance matrix time series model; Profiled quasi maximum likelihood estimation; Realized covariance matrix; Semiparametric time series model; Structural change testing.

ing multivariate process (see, e.g., [Barndorff-Nielsen and Shephard \(2002, 2004\)](#) and [Andersen et al. \(2003\)](#)). Since the modelling and forecast of 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 RCOV matrix is its positive definiteness. To preserve this feature, many covariance matrix time series models were proposed based on the Wishart or matrix-F innovation distribution, which can generate random positive definite matrices automatically. The related works in this direction include [Gouriéroux et al. \(2009\)](#), [Golosnoy et al. \(2012\)](#), [Yu et al. \(2017\)](#), [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 change in applications; see, e.g., [Aue and Horváth \(2013\)](#) for an overview. In the presence of structural change, the conventional statistical methodologies could lead to a severe deterioration in predictions ([Pesaran and Timmermann, 2004](#)), and at the same time, some 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 spurious long memory phenomenon caused by structural change. This indicates that the existing long memory models ([Chiriac and Voev, 2011](#); [Corsi, 2009](#)) for the RCOV matrix data should be used with caution, since they have overlooked the possibility of spurious long memory. To circumvent the aforementioned dilemmas, a formal test for detecting the structural change in matrix time series models is needed, but has not been attempted before.

This paper aims to propose a new generalized Hausman test for detecting the 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 the abrupt and smooth structural change with an unknown form in the spirit of [Robinson \(1989, 1991\)](#). The generalized Hausman testing framework was initiated by [Chen and Hong \(2012\)](#), and it was 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 seems the first formal attempt to detect the structural change for matrix time series models. Under suitable conditions, our test is shown to have a standard normal limiting null distribution, and also non-trivial power in detecting a broad class of abrupt and smooth local alternatives.

We shall mention that testing for structural change is a long-standing problem in the literature. Besides the generalized Hausman test, many other tests have been studied for regression or univariate/multivariate time series models; see, e.g., the cumulative sum type tests in [Brown et al. \(1975\)](#), [Hidalgo \(1995\)](#), [Su and Xiao \(2008\)](#), [Aue et al. \(2009\)](#), and [Shao and Zhang \(2010\)](#); the supremum type tests in [Andrews \(1993\)](#), [Davis et al. \(1995\)](#), [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, since it has several appealing advantages over the other tests. First, it can detect not only abrupt changes but also smooth changes without assuming a specified structural change alternative or knowing the type of changes. Clearly, it is important for the test to avoid such potential model mis-specification. 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, the test thereby possesses the power to detect changes happened very early or late.

Once structural change is indicated by our generalized Hausman test, it is important to have a dynamic model to capture this structural change. For this purpose, we further propose a new semiparametric covariance matrix time series model in this paper. Our semiparametric model has a multiplicative form of combining a time-varying deterministic long run component and a BEKK-type (Engle and Kroner, 1995) short run component. Hence, it is termed the time-varying BEKK (TVB) model. For this non-stationary 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 by a profiled quasi maximum likelihood estimator (QMLE). This profiled QMLE is obtained by assuming that the model innovation follows the Wishart distribution, and it 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 changes 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, and this requires non-trivial proof techniques, since 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 one real 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.

Some notations are 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. Also, 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 the convergence in distribution, and the range of integration is $[-1, 1]$ unless stated otherwise.

2. Testing for Structural Change.

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 $\mathbf{y}_t$:

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

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 un-specified 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 pioneer works in Robinson (1989, 1991) for 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, we highlight that the possible non-stationarity of $\mathbf{y}_t$ in model (2.1) is caused by the deterministic Σ_t , and this setting is in accordance with the semiparametric model in Section 3 below but does not allow Σ_t to be non-stationary driven by some unit root type process. To study the unit root type non-stationarity 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 RCOV matrix, and such approach implicitly assumes the stationarity of $\mathbf{y}_t$, or equivalently, the constancy of Σ_t . However, the reality abounds with changes, and in fact, they are the rule rather than the exception. It is well documented that ignoring structural changes could mislead the conventional time series analysis procedure, resulting in erroneous conclusions (Pesaran and Timmermann, 2004). This motivates us to propose a formal structural change test for model (2.1). Our another motivation for the structural change testing is triggered by the spurious long memory phenomenon (see more numerical evidences in the Supplementary Material). However, this important aspect has been overlooked in matrix time series applications. For instance, the realized covariance (RCOV) matrices time series data are often observed having the long memory phenomenon, and empirical researchers usually apply long memory models to fit the RCOV matrices

data (Chiriac and Voev, 2011; Corsi, 2009). We argue that these fitted long memory models should be used with caution, since they do not take the possibility of spurious long memory phenomena into account. 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 it allows the structure of $\mathbf{y}_t$ to change smoothly or abruptly. Testing for structural change has been well studied in regression or conditional mean time series models; see Fu and Hong (2019) and references therein. In conditional variance models, most of the efforts are made for GARCH models; see Berkes et al. (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 the multivariate conditional variance models with unknown form, the related structural change tests are absent, and this is also true for our matrix model (2.1) with $n > 1$.

2.2. Test statistic. In this paper, we adopt the idea of generalized Hausman test in Chen and Hong (2012) to propose a structural change test for model (2.1). To elaborate this new test, 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 can be proposed by measuring the distance of these two estimators under certain norm to look for the 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. Since $\frac{1}{T} \sum_{s=1}^T K_h(x - s/T) = 1 + O(T^{-1}h^{-1})$ by the integrability of $K(\cdot)$, it is more convenient to consider an alternative estimator of $\Sigma(x)$ given by

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

In view of 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, i.e., there is insufficient information for estimators at the extreme left and right ends of the sample interval. To avoid such a problem, following [Chen and Hong \(2012\)](#), we can 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 as in the interior region $[Th] < t < T-[Th]$, and modify $\hat{\Sigma}_t$ as

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

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

Another way to circumvent the edge effect is using 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) can still apply to the test statistic defined in (2.4) below with some appropriate modifications. However, using the local linear estimator will complicate the estimation theory for the proposed semiparametric model in Section 3 below. Hence, we only focus on the kernel (local constant) estimator for brevity in this paper, and leave the investigation of the local linear estimator for future study.

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

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

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

$$(2.4) \quad \hat{\mathcal{D}} = \frac{Th^{1/2}\hat{\mathcal{S}} - \hat{\mathcal{B}}}{\hat{\mathcal{V}}^{1/2}},$$

where $\hat{\mathcal{B}} = h^{-1/2} \text{tr}(\hat{M}) \left[\int K^2(x) dx \right]$, $\hat{\mathcal{V}} = 2 \text{tr}(\hat{M}^2) \int \left[\int K(x) K(x + \lambda) dx \right]^2 d\lambda$, 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 $\hat{\mathcal{B}}$ and $\hat{\mathcal{V}}$ are the estimators of the mean and the variance of $Th^{1/2}\hat{\mathcal{S}}$, respectively. 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 Andrews (1991); Newey and West (1987)). We shall highlight that the test in Chen and Hong (2012) does not need to consider the corresponding long run variance estimator, since 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 variable. 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 $\hat{\mathcal{D}}$ 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 $\hat{\mathcal{D}}$, the following four assumptions are needed.

ASSUMPTION 2.1. $\Sigma(x)$ 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 2.2. (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 2.3. (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 2.4. (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 $\lambda_h/2 < \lambda_b < 1/2 - 1/2\lambda_h$ for some $0 < \lambda_h < 1/2$.

Assumption 2.1 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 2.2–2.4 are in line with Fu and Hong (2019). Assumption 2.2(i) requires some sufficient technical conditions on $\mathbf{u}_t$. The stationarity condition for $\mathbf{u}_t$ might be restrictive under some circumstances. A desired future study 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 3.1 below, Assumption 2.2(i) holds when $\mathbf{u}_t$ follows model (3.1) with $\rho(\mathcal{A}_0 + \mathcal{B}_0) < 1$. Assumption 2.2(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

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

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

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

with $z_t = \text{vech}(\mathbf{u}_t - I_n)$. Assumption 2.3(i) holds for commonly used kernels such as uniform, Epanechnikov, truncated Gaussian, and many others. Assumption 2.3(ii) requires that h converges to zero at a slower rate than $T^{-1/2}$. Assumption 2.4(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 be $[-1, 1]$. Assumption 2.4(ii) gives a sufficient technical condition

on b_T to ensure the consistency of $\widehat{M}$ under the null and it was adopted in [Hong et al. \(2017\)](#) and [Fu and Hong \(2019\)](#).

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

THEOREM 2.1. *Suppose Assumptions 2.1–2.4 hold. Then, under H_0 ,*

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

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

$$(2.7) \quad \widehat{\mathcal{D}} > c_\alpha,$$

which has an asymptotic significance level α by Theorem 2.1.

To compute $\widehat{\mathcal{D}}$, we need 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 by choosing $h = T^{-1/5}/\sqrt{12}$ as in [Fu and Hong \(2019\)](#) works well, where 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 could be optimally chosen by minimizing the integrated mean square error ([Robinson, 1989](#)), however, the resulting chosen c_h may not be optimal for the test as pointed out by [Hong et al. \(2017\)](#). In general, to obtain the optimal bandwidth for $\widehat{\mathcal{D}}$, we may need higher order Edgeworth expansions for the finite sample distribution of $\widehat{\mathcal{D}}$, and we leave this interesting topic for future study.

To select kernel function k and the truncated lag b_T , we use the method in [Andrews \(1991\)](#) so that the mean and variance of $Th^{1/2}\widehat{\mathcal{S}}$ 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, since 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 could 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

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

(see (6.5) in [Andrews \(1991\)](#)), where $(\widehat{\phi}_a, \widehat{\psi}_a)$ is the least squares estimator of (ϕ, ψ) in the ARMA(1, 1) model: $\bar{y}_{a,t} = \phi \bar{y}_{a,t-1} + \psi \tau_{a,t-1} + \tau_{a,t}$, $\widehat{\sigma}_a^2$ 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 $\widehat{\pi}$ performs well in finite samples.

Next, we study the power behavior of our test $\widehat{\mathcal{D}}$ under the local alternative:

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

where $\Sigma_{1t} = \Sigma_1(t/T)$, and the deterministic matrix function $\Sigma_1(x)$ satisfies the conditions for $\Sigma(x)$ in Assumption 2.1, 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 $\widehat{\mathcal{S}}$ using the Frobenius norm can be viewed as a L_2 type statistic, and its convergence rate in Theorem 2.1 is $Th^{1/2}$. It 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 non-trivial local power of $\widehat{\mathcal{D}}$.

THEOREM 2.2. *Suppose Assumptions 2.1–2.3, 2.4(i) hold, and Assumption 2.4(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 $\frac{T^{1/2}h^{1/4}}{c_T} \rightarrow 0$, then $\widehat{\mathcal{D}} \rightarrow_{\mathcal{L}} N(0, 1)$;*
- (ii) if $\frac{T^{1/2}h^{1/4}}{c_T} \rightarrow c \in (0, \infty)$, then $\widehat{\mathcal{D}} \rightarrow_{\mathcal{L}} N(c^2\mathcal{B}_v, 1)$;*
- (iii) if $\frac{T^{1/2}h^{1/4}}{c_T} \rightarrow \infty$, then $\widehat{\mathcal{D}} \rightarrow \infty$ in probability;*

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

The above theorem implies that $\widehat{\mathcal{D}}$ 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 $h \propto T^{-1/5}$, $T^{1/2}h^{1/4} \propto T^{9/20}$ is close to $T^{1/2}$. The same local rate $T^{1/2}h^{1/4}$ has also been considered in Härdle and Mammen (1993), Chen and Hong (2012), Fu and Hong (2019), and many others.

Moreover, the modification of Assumption 2.4(ii) is tailored to give a sufficient technical condition on b_T to ensure the consistency of $\widehat{M}$ for Theorem 2.2(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 2.2.

3. A Semiparametric Matrix Model.

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

$$(3.1) \quad \mathbf{u}_t = G_t^{1/2} \mathbf{e}_t G_t^{1/2}$$

$$\text{with } G_t = I_n - A_0 A_0' - B_0 B_0' + A_0 \mathbf{u}_{t-1} A_0' + B_0 G_{t-1} B_0',$$

where $G_t \in \mathcal{F}_{t-1}$ has a first order BEKK specification in Engle and Kroner (1995) with $EG_t = I_n$, and $\{\mathbf{e}_t\}_{t=1}^T$ is a sequence of i.i.d. positive definite random innovation matrices with $E\mathbf{e}_t = I_n$. With minor modifications, our studies can be easily extended to the case that G_t has a higher order mean-targeted BEKK specification, which however will not be considered 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 initiated by the univariate volatility model in Feng (2004), and it has been becoming popular to describe the dynamic of the univariate volatility (see, e.g., Engle and Rangel (2008); Xu and Phillips (2008); Chen and Hong (2012); Zhu (2019); Zhang et al. (2020); Jiang et al. (2021)) as well as the multivariate volatility (see, e.g., Hafner and Linton (2010); Patilea and Raïssi (2012); Amado and Teräsvirta (2014)). The TVB model can thus be viewed as a matrix generalization of the model in Feng (2004). Also, 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 $\hat{\mathbf{y}}_t(s)$ 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. (2021), we have the following proposition, which gives the stationarity and moment properties of $\mathbf{u}_t$.

PROPOSITION 3.1. *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 3.1 and the boundedness of $\Sigma(x)$. If we do not require $E \|\mathbf{u}_t\| < \infty$, we may have 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, e.g., 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 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 $\mathbf{y}_t$ below, and the details are straightforward and hence omitted.

PROPOSITION 3.2. *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 estimation 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 3.1. $\mathbf{u}_t$ is strictly stationary and ergodic with $E \|\mathbf{u}_t\|^2 < \infty$.

ASSUMPTION 3.2. $h = c_h T^{-\lambda_h}$ for some $1/5 \leq \lambda_h < 1$ and $0 < c_h < \infty$.

Assumption 3.1 gives some regularity conditions on $\mathbf{u}_t$, which imply that $\mathbf{u}_t$ is β -mixing with exponential decay by Proposition 3.1. Assumption 3.2 requires that h converges to zero at a slower rate than T^{-1} , and it is weaker than Assumption 2.3(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

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

With these notations, the asymptotics of $\hat{\Sigma}(x)$ are given in the following theorem.

THEOREM 3.1. Suppose Assumptions 2.1, 2.3(i) and 3.1–3.2 hold. Then, for any $x \in (0, 1)$,

(i) $\hat{\sigma}(x) \rightarrow \frac{1}{2}[\sigma(x-) + \sigma(x+)]$ 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

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

REMARK 1. At continuous points, Theorem 3.1 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 3.2 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 estimation procedure next estimates ϕ via a profiled quasi maximum likelihood estimator (QMLE). Specifically, we write the parametric G_t as

$$(3.2) \quad G_t(\phi) = I_n - AA' - BB' + A\mathbf{u}_{t-1}A' + BG_{t-1}(\phi)B'$$

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

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

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, since our main interest is to estimate ϕ .

The reason that we consider the Wishart log-likelihood function in estimation is two-fold: 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 et al., 2021), the Wishart profiled QMLE of ϕ_0 below does not require any distributional assumption on $\mathbf{e}_t$ for its asymptotics.

Due 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 $\hat{\Sigma}_t$ and $\hat{\mathbf{u}}_t$, respectively, where $\hat{\mathbf{u}}_t = \hat{\Sigma}_t^{-1/2} \mathbf{y}_t \hat{\Sigma}_t^{-1/2}$. Consequently, we consider the following profiled log-likelihood function:

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

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

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

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

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

$$(3.6) \quad \hat{\gamma} = \arg \min_{\gamma \in \Gamma} \hat{L}_T(\gamma),$$

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 $\hat{\phi}$ is irrelevant to ν , leading to

$$(3.7) \quad \hat{\phi} = \arg \min_{\phi \in \Gamma_\phi} \widehat{\mathcal{L}}_T(\phi).$$

Although $\hat{\phi}$ is defined by assuming $\mathbf{e}_t \sim \nu^{-1} \text{Wishart}(\nu, I_n)$, the consistency and asymptotic normality of $\hat{\phi}$ hold for any distribution of $\mathbf{e}_t$ with $E\mathbf{e}_t = I_n$ (see Theorem 3.2 below). Hence, we call $\hat{\phi}$ the profiled QMLE of ϕ_0 .

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

ASSUMPTION 3.3. (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 3.4. Γ_ϕ is a compact subset of $\mathbb{R}^{2n^2}$, and ϕ_0 is an interior point of Γ_ϕ .

ASSUMPTION 3.5. $h = c_h T^{-\lambda_h}$ for some $1/4 < \lambda_h < 1/2$ and $0 < c_h < \infty$.

Assumption 3.3(i) rules out the deterministic situation of the BEKK process, Assumption 3.3(ii) ensures the positive definiteness of $G_t(\phi)$, and Assumption 3.3(iii) is standard for model identifiability; see Engle and Kroner (1995). Assumption 3.4 gives some regularity conditions for studying the model estimation. Assumption 3.5 is the key to derive the asymptotics of $\hat{\phi}$. A restricted h is required to undersmooth the estimation of Σ_t so that the impact of its bias to the asymptotics of $\hat{\phi}$ can be neglected; see Hafner and Linton (2010), Jiang et al. (2021), and Remark 4 below for more discussions.

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

$E(\mathbf{u}_{t-j} \otimes G_t^{-1} \frac{\partial G_t}{\partial \phi_i} G_t^{-1})$, and

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

With these notations, 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

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

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, respectively, according to

$$\text{vec}(N) = E(\rho_t^{\otimes 2}) \text{vec}(\text{Var}(\xi_t)),$$

$$\text{vec}(\Psi) = \left[\int_0^1 \Upsilon(x)^{\otimes 2} dx \right] \Xi_0^{\otimes 2} E[(G_t^{1/2})^{\otimes 4}] \text{vec}(\text{Var}(\xi_t)),$$

$$\text{vec}(H) = \left[\int_0^1 (\Upsilon(x) \otimes I_{2n^2}) dx \right] [\Xi_0 \otimes I_{2n^2}] E[(G_t^{1/2})^{\otimes 2} \otimes \rho_t] \text{vec}(\text{Var}(\xi_t))$$

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 3.2. *Suppose Assumptions 2.1, 2.3(i) and 3.1–3.4 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 3.2 is replaced by Assumption 3.5,

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

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 $\Sigma(x)$, since its asymptotic variance is invariant regardless of the form of $\Sigma(x)$. However, when $n > 1$, $\hat{\phi}$ is not adaptive any more caused by 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 et al. \(2021\)](#). Therefore, when $n > 1$, if there exists cross-sectional dependence between any two elements of $\mathbf{y}_t$, the estimation of Σ_t should have the impact on the asymptotic distribution of $\hat{\phi}$. Note that when $n = 1$, Σ_t only represents the scale of $\mathbf{y}_t$ and has no effect on the dynamic of $\mathbf{u}_t$, and this may be the reason for the adaptiveness of $\hat{\phi}$ in this univariate case.

REMARK 3. To prove Theorem 3.2(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, simulation studies in Section 4 show that the asymptotic normality of $\hat{\phi}$ in Theorem 3.2(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 at the moment and hence the topic is left for future work.

REMARK 4. To obtain the result in Theorem 3.2(ii), we need Assumption 3.5 for our undersmoothing technique, which makes the bias from the first step non-parametric estimator $\hat{\Sigma}_t$ having a negligible effect on the asymptotics of $\hat{\phi}$. Due to Assumption 3.5, the popular choice of bandwidth $h = c_h T^{-1/5}$ is excluded. To relax Assumption 3.5, one may use the bias correction and small bandwidth techniques (see, e.g., [Newey et al. \(2004\)](#), [Aradillas-Lopez et al. \(2007\)](#), [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 some future investigations.

In practice, we can estimate V_{ϕ_0} consistently by its sample version. Specifically, let $\hat{G}_t = \hat{G}_t(\hat{\phi})$, $\frac{\partial \hat{G}_t}{\partial \phi_i} = \frac{\partial \hat{G}_t(\hat{\phi})}{\partial \phi_i}$, $\hat{\mathbf{e}}_t = \hat{G}_t^{-1/2} \hat{\mathbf{u}}_t \hat{G}_t^{-1/2}$ 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 $\hat{J} = \frac{1}{T} \sum_{t=1}^T \hat{\rho}_t \hat{\rho}_t'$ and $\hat{M}_j^i = \frac{1}{T} \sum_{t=j+1}^T \hat{\mathbf{u}}_{t-j} \otimes \hat{G}_t^{-1} \frac{\partial \hat{G}_t}{\partial \phi_i} \hat{G}_t^{-1}$, respectively. To compute the sample version of Q_{ϕ_0} , we first let

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

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 2.4(ii). Based on $\hat{\Upsilon}(t/T)$, we next compute the sample versions of N , Ψ and H , respectively, by

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

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

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

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 $\widehat{W}$ is larger than the upper-tailed critical value of χ_s^2 .

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

$$(3.10) \quad \widehat{\nu} = \arg \min_{\nu \in \Gamma_\nu} \widehat{L}_T(\widehat{\phi}, \nu).$$

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

THEOREM 3.3. *Suppose $\mathbf{e}_t \sim \nu_0^{-1} \text{Wishart}(I_n, \nu_0)$ and the conditions in Theorem 3.2 hold. Then,*

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

(ii) $\sqrt{T}(\widehat{\nu} - \nu_0) \rightarrow_{\mathcal{L}} N(0, 2J_{\nu_0}^{-1})$ as $T \rightarrow \infty$, where $J_{\nu_0} = \frac{1}{2} \sum_{i=1}^n \psi'(\frac{\nu_0+1-i}{2}) - \frac{n}{\nu_0}$,

and $\psi(\cdot)$ is the digamma function.

By some simple calculations, we can see that when $\mathbf{e}_t \sim \nu_0^{-1} \text{Wishart}(I_n, \nu_0)$, the Cramér-Rao lower bound with respect to ν_0 is $2J_{\nu_0}^{-1}$. Interestingly, Theorem 3.3(ii) shows that $\widehat{\nu}$ 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 via some simulation experiments.

4.1. *Study on the test.* In this subsection, we examine the finite sample performance of the generalized Hausman test $\widehat{D}$. We generate 1000 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:

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

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

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

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

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

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

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

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

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 $F(\nu_0^\dagger, \frac{\nu_{20} - n - 1}{\nu_{10}} I_n)$ 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 $\widehat{\mathcal{D}}$ according to statements in Subsection 2.3

Specifically, we consider three choices of the 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; meanwhile, 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 $b_T = 1.3221(\widehat{\pi}T)^{1/5}$, where $\widehat{\pi} = \min(1600, \widehat{\pi}_2)$ and $\widehat{\pi}_2$ is defined in (2.8). When the value of $\widehat{\mathcal{D}}$ falls into the critical region in (2.7), we reject H_0 at level α .

Because the size performance of $\widehat{\mathcal{D}}$ may be a bit sensitive to the choice of h , we also follow the suggestion in Fu and Hong (2019) to construct a bootstrapped critical region and make decisions via 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

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

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 $l_T^B = \lfloor \frac{3}{2}c_B^2 T \rfloor^{1/3}$, where $c_B = \text{tr}(\sum_{j=-b_T}^{b_T} k(\frac{j}{b_T})|j|\hat{\Gamma}_{v,j})/\text{tr}(\hat{M})$, and $\hat{M}$ and $\hat{\Gamma}_{v,j}$ are defined as in (2.4). Currently, we are unable to provide the consistency for the above bootstrap method, although our simulation results below indicate that this bootstrap method works well. The reason is because the existing techniques for proving bootstrap consistency in parametric models (see, e.g., Künsch (1989)) are not directly transferrable to deal with the current setting with the presence of the nonparametric kernel estimation.

Table 1 reports the sizes of $\hat{\mathcal{D}}$, where $\hat{\mathcal{D}}^{AC}$ and $\hat{\mathcal{D}}^{BC}$ 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, $\hat{\mathcal{D}}^{AC}$ is undersized, but the size of $\hat{\mathcal{D}}^{BC}$ is more accurate; (ii) both $\hat{\mathcal{D}}^{AC}$ and $\hat{\mathcal{D}}^{BC}$ have satisfactory size performances regardless of the choice of $K(x)$ and the distribution of $\mathbf{e}_t$ when the sample size T increases, and $\hat{\mathcal{D}}^{BC}$ has a slightly better size performance than $\hat{\mathcal{D}}^{AC}$ especially at level 10%; (iii) the size of $\hat{\mathcal{D}}^{BC}$ 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 $\hat{\mathcal{D}}$ under the alternative settings in (4.2)–(4.5). As before, we consider both $\hat{\mathcal{D}}^{AC}$ and $\hat{\mathcal{D}}^{BC}$. Their results at level 5% are reported in Fig 1 for the case that the data are generated from DGP 1, $T = 1500$ and $T = 3000$ with $\mathbf{e}_t \sim \text{Wishart}$ and $\hat{\mathcal{D}}$ is calculated by choosing $K(x) \sim U$. The results for other cases are similar, and hence they are not reported for saving space. From Fig 1, we find that (i) $\hat{\mathcal{D}}^{AC}$ is slightly more powerful than $\hat{\mathcal{D}}^{BC}$; (ii) the power of $\hat{\mathcal{D}}^{AC}$ and $\hat{\mathcal{D}}^{BC}$ always increases with r as expected and reaches 1 for large r ; (iii) $\hat{\mathcal{D}}^{AC}$ and $\hat{\mathcal{D}}^{BC}$ under the off diagonal smooth (or multiple abrupt) change alternative are generally

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

		$\mathbf{e}_t \sim \text{Wishart}$				$\mathbf{e}_t \sim \text{matrix-F}$			
		$\widehat{\mathcal{D}}^{AC}$		$\widehat{\mathcal{D}}^{BC}$		$\widehat{\mathcal{D}}^{AC}$		$\widehat{\mathcal{D}}^{BC}$	
T	$K(x)$	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
1000	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
1500	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
3000	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
1000	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
1500	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
3000	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

more powerful than the diagonal smooth (or single abrupt) change alternative for any fixed r .

Overall, $\widehat{\mathcal{D}}^{AC}$ is better than $\widehat{\mathcal{D}}^{BC}$ in terms of the power to detect both smooth and abrupt changes, while $\widehat{\mathcal{D}}^{BC}$ shows some advantage over $\widehat{\mathcal{D}}^{AC}$ in terms of the size. Considering the tradeoff of size accuracy and power enhancement, we recommend to look at the results of both $\widehat{\mathcal{D}}^{AC}$ and $\widehat{\mathcal{D}}^{BC}$ 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 1000 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 $r = 10$.

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

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 Uniform kernel function for $K(x)$, since the results based on other choices of $K(x)$ are similar; and we do not take $h \propto T^{-1/5}$ suggested by the conventional cross-validation method, as it violates Assumption 3.5.

Tables 2 and 3 report the sample bias, the sample empirical standard deviation (ESD) and the average asymptotic standard deviation (ASD) of $\widehat{\phi}$ based on DGPs 1 and 2, respectively, where the ASD is calculated according to Theorem 3.2(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 corresponding values of ESD (or ASD) in the case of $\mathbf{e}_t \sim \text{matrix-F}$, and this is consistent with the fact that $\hat{\phi}$ is an efficient estimator when $\mathbf{e}_t$ is the Wishart distributed. Note that when $\mathbf{e}_t \sim \text{Wishart}$, Tables 2 and 3 also give the results for the estimate $\hat{\nu}$, which shows good performance as $\hat{\phi}$.

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

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	1500	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
	3000	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	1500	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	\
	3000	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	1500	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
	3000	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	1500	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	\
	3000	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	\

5. An Empirical Example. In this section, we revisit the RCOV matrix time series data of IBM Common Stock (IBM) and Microsoft Corporation (MSFT) in Lunde et al. (2016). This data set ranges from January 2006 to December 2011 with 1474 observations in total, and two flash crashes in 6 May, 2010 and 9 August, 2011 are replaced by the average of their nearest five preceding and following matrices,

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	1500	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
	3000	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	1500	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	\
	3000	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	1500	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
	3000	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	1500	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	\
	3000	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	\

respectively. Denote this data set by $\{\mathbf{y}_t\}_{t=1}^{1474}$. First, we plot $\{\mathbf{y}_{rs,t}\}_{t=1}^{1474}$ as well as the sample ACFs of $\mathbf{y}_{rs,t}$ in Fig 2, where $\mathbf{y}_{rs,t}$ is (r, s) th element of $\mathbf{y}_t$. From this figure, we can witness a peak in all $\mathbf{y}_{rs,t}$ around the late 2008, and the sample ACFs of each $\mathbf{y}_{rs,t}$ decay slowly.

Next, we apply our generalized Hausman test $\hat{\mathcal{D}}$ to detect whether there is any structural change in $\mathbf{y}_t$. Based on this data set, the value of $\hat{\mathcal{D}}$ is 4.2061, where $\hat{\mathcal{D}}$ is computed as in Subsection 4.1 with $K(x) \sim U$. Moreover, the p-value of $\hat{\mathcal{D}}$ using asymptotic critical region (as for $\hat{\mathcal{D}}^{AC}$) and bootstrapped critical region (as for $\hat{\mathcal{D}}^{BC}$) is 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 $\mathbf{y}_t$ by our semiparametric TVB model, which is estimated by the two-step estimation procedure in Subsection 3.3. The top panels of Fig 3 plot the val-

ues of $\hat{\Sigma}_{rs,t}$ (the (r, s) th element of $\hat{\Sigma}_t$), which are computed as in Subsection 4.2. In view of $\hat{\Sigma}_{rs,t}$, 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 the other time period. This finding is reasonable, since the subprime financial crisis happened in 2008-2009, and the (co-)variance of IBM and MSFT tends to be higher during that period. With 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 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 a strong evidence to reject H''_0 but not H'_0 . Consequently, we then re-fit the short run component G_t by 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 $\{\hat{\mathbf{e}}_{rs,t}\}$ (denoted by $\rho_{rs}(j)$ for $j = 1, \dots, 100$) in three bottom panels of Fig 3, where $\hat{\mathbf{e}}_{rs,t}$ 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 2 or 3 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 some efforts could be further made to capture the possible dependence within $\mathbf{e}_{12,t}$ and $\mathbf{e}_{22,t}$ by using high-order or long-memory TVB models.

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

FIG 3. Top panels: $\hat{\Sigma}_{rs,t}$. Bottom panels: sample ACFs of $\mathbf{e}_{rs,t}$.

The conditional heterogeneous autoregressive (HAR) model proposed by [Corsi \(2009\)](#) is a benchmark to study the univariate RCOV, since it can capture the observed

long memory feature of RCOV. As shown in Fig 2, our considered RCOV matrix data also exhibit the long memory feature, and they thus can be fitted by an alternative stationary HAR-type matrix time series model:

$$(5.1) \quad \mathbf{y}_t = G_t^{1/2} \mathbf{e}_t G_t^{1/2},$$

$$G_t = \Omega + A_{(d)} \mathbf{y}_{t-1,d} A'_{(d)} + A_{(w)} \mathbf{y}_{t-1,w} A'_{(w)} + A_{(m)} \mathbf{y}_{t-1,m} A'_{(m)},$$

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 of RCOV matrices, respectively. 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 compare the forecasting performance of our TVB model with this HAR model, based on a rolling window procedure with window size equal to 1000. 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 the smaller forecast error than the HAR model. Moreover, by using 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 a significant forecasting accuracy over the HAR model. This forecasting advantage of the TVB model over the HAR model may be due to its ability to take

the structural change of $\mathbf{y}_t$ into account.

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

6. Concluding Remarks. This paper first provides a generalized Hausman test to detect the structural change in covariance matrix time series, and derives its limiting distributions under both null and local alternatives. As illustrated by simulations, the 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 it is applicable without assuming any prior information on the structural change alternative. This merit makes our test appealing, since it has no guarantee that the structures of matrix time series specified by the researcher provide a correct description of reality.

Moreover, this paper proposes a new semiparametric TVB model, which simultaneously allows for structural change as well as temporal dependence. Our TVB model is estimated by a two-step estimation procedure, and the asymptotics of its related estimators are established. Since this two-step estimation procedure is valid without specifying the form of structural change or the distribution of innovation, it can have 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 structural change over a BEKK-type model during the examined period, and its observed long memory phenomenon is well captured by the TVB model. Of course, the structural change may not be the only reason to cause long-memory phenomenon, and 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 the study on 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. How to extend our work to the high dimensional setting (see, e.g., [Tao et al. \(2011\)](#), [Leng and Tang \(2012\)](#) and [Wang et al. \(2019\)](#)) could be another promising direction for future research.

Acknowledgements. 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 the 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).

Supplementary Material. The supplement to this paper contains Appendices A, B, C and D, where Appendix A gives the proofs of Theorems [2.1–2.2](#), Appendix B offers the proofs of Theorems [3.1–3.3](#), Appendix C lists some basic derivatives results, and Appendix D provides some numerical evidences on spurious long memory phenomena caused by the structural change.

References.

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

- [2] Amado, C. and Teräsvirta, T. (2014). Conditional correlation models of autoregressive conditional heteroscedasticity with nonstationary GARCH equations. *Journal of Business & Economic Statistics*, 32(1):69–87.
- [3] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.
- [4] Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.
- [5] Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856.
- [6] Aradillas-Lopez, A., Honoré, B. E., and Powell, J. L. (2007). Pairwise difference estimation with nonparametric control variables. *International Economic Review*, 48(4):1119–1158.
- [7] Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *Annals of Statistics*, 37(6B):4046–4087.
- [8] Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16.
- [9] Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.
- [10] 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(2):253–280.
- [11] Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.
- [12] Berkes, I., Horváth, L., and Kokoszka, P. (2004). Testing for parameter constancy in GARCH(p, q) models. *Statistics & Probability Letters*, 70(4):263–273.
- [13] 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(2):149–163.
- [14] Cattaneo, M. D. and Jansson, M. (2018). Kernel-based semiparametric estimators: small bandwidth asymptotics and bootstrap consistency. *Econometrica*, 86(3):955–995.

- [15] Cavaliere, G. and Taylor, A. R. (2007). Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics*, 140(2):919–947.
- [16] Chan, N. and Wang, Q. (2015). Nonlinear regressions with nonstationary time series. *Journal of Econometrics*, 185(1):182–195.
- [17] Chen, B. (2015). Modeling and testing smooth structural changes with endogenous regressors. *Journal of Econometrics*, 185(1):196–215.
- [18] Chen, B. and Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80(3):1157–1183.
- [19] Chen, B. and Hong, Y. (2016). Detecting for smooth structural changes in GARCH models. *Econometric Theory*, 32(3):740–791.
- [20] Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947.
- [21] Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.
- [22] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1):1–37.
- [23] Davis, R. A., Huang, D., and Yao, Y.-C. (1995). Testing for a change in the parameter values and order of an autoregressive model. *Annals of Statistics*, 23(1):282–304.
- [24] Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1):134–144.
- [25] Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150.
- [26] 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(3):1187–1222.
- [27] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*, volume 66. Chapman & Hall, London.
- [28] Fan, J. and Yao, Q. (2008). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.
- [29] Feng, Y. (2004). Simultaneously modeling conditional heteroskedasticity and scale change. *Econometric Theory*, 20(3):563–596.

- [30] 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)*, 5(76):903–924.
- [31] Fu, Z. and Hong, Y. (2019). A model-free consistent test for structural change in regression possibly with endogeneity. *Journal of Econometrics*, 211(1):206–242.
- [32] Golosnoy, V., Gribisch, B., and Liesenfeld, R. (2012). The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics*, 167(1):211–223.
- [33] Gouriéroux, C., Jasiak, J., and Sufana, R. (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2):167–181.
- [34] Hafner, C. M. and Linton, O. (2010). Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics*, 159(1):55–73.
- [35] Hafner, C. M. and Preminger, A. (2009). On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis*, 100(9):2044–2054.
- [36] 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(415):665–672.
- [37] Hansen, B. E. (2000). Testing for structural change in conditional models. *Journal of Econometrics*, 97(1):93–115.
- [38] Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, 21(4):1926–1947.
- [39] Hidalgo, J. (1995). A nonparametric conditional moment test for structural stability. *Econometric Theory*, 11(4):671–698.
- [40] Hong, Y., Wang, X., and Wang, S. (2017). Testing strict stationarity with applications to macroeconomic time series. *International Economic Review*, 58(4):1227–1277.
- [41] Jiang, F., Li, D., and Zhu, K. (2021). Adaptive inference for a semiparametric generalized autoregressive conditional heteroskedasticity model. *Journal of Econometrics*, In Press.
- [42] Konno, Y. (1991). A note on estimating eigenvalues of scale matrix of the multivariate f -distribution. *Annals of the Institute of Statistical Mathematics*, 43(1):157–165.
- [43] Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of statistics*, 17(3):1217–1241.

- [44] Leng, C. and Tang, C. Y. (2012). Sparse matrix graphical models. *Journal of the American Statistical Association*, 107(499):1187–1200.
- [45] Lima, L. R. and Xiao, Z. (2010). Is there long memory in financial time series? *Applied Financial Economics*, 20(6):487–500.
- [46] Lin, C.-F. J. and Teräsvirta, T. (1994). Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics*, 62(2):211–228.
- [47] 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(4):504–518.
- [48] 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(1):378–390.
- [49] Newey, W. K., Hsieh, F., and Robins, J. M. (2004). Twicing kernels and a small bias property of semiparametric estimators. *Econometrica*, 72(3):947–962.
- [50] Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.
- [51] 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(4):643–657.
- [52] Patilea, V. and Raïssi, H. (2012). Adaptive estimation of vector autoregressive models with time-varying variance: application to testing linear causality in mean. *Journal of Statistical Planning and Inference*, 142(11):2891–2912.
- [53] 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(3):411–425.
- [54] Politis, D. N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70.
- [55] Qu, Z. and Perron, P. (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica*, 75(2):459–502.
- [56] Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change*, pages 253–264. Springer.
- [57] Robinson, P. M. (1991). Time-varying nonlinear regression. In *Economic Structural Change*, pages 179–190. Springer.

- [58] Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, 105(491):1228–1240.
- [59] Stărică, C. and Granger, C. (2005). Nonstationarities in stock returns. *Review of Economics and Statistics*, 87(3):503–522.
- [60] Su, L. and Xiao, Z. (2008). Testing structural change in time-series nonparametric regression models. *Statistics and its Interface*, 1(2):347–366.
- [61] 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(495):1025–1040.
- [62] Vogt, M. (2015). Testing for structural change in time-varying nonparametric regression models. *Econometric Theory*, 31(4):811–859.
- [63] Wang, D., Liu, X., and Chen, R. (2019). Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics*, 208(1):231–248.
- [64] Wang, Q. and Phillips, P. C. B. (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics*, 40(2):727–758.
- [65] Xu, K.-L. and Phillips, P. C. B. (2008). Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics*, 142(1):265–280.
- [66] 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(4):513–527.
- [67] Zhang, T. and Wu, W. B. (2012). Inference of time-varying regression models. *Annals of Statistics*, 40(3):1376–1402.
- [68] Zhang, Y., Liu, R., Shao, Q., and Yang, L. (2020). Two-step estimation for time varying ARCH models. *Journal of Time Series Analysis*, 41(4):551–570.
- [69] Zhou, J., Jiang, F., Zhu, K., and Li, W. K. (2021). Time series models for realized covariance matrices based on the matrix-F distribution. *Statistica Sinica*, In Press.
- [70] 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(4):513–531.
- [71] Zhu, K. (2019). Statistical inference for autoregressive models under heteroscedasticity of unknown form. *Annals of Statistics*, 47(6):3185–3215.