<table>
<thead>
<tr>
<th><strong>Statistica Sinica Preprint No:</strong> SS-2021-0029</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Manuscript ID</strong></td>
</tr>
<tr>
<td><strong>URL</strong></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
</tr>
<tr>
<td><strong>Complete List of Authors</strong></td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
</tr>
<tr>
<td><strong>E-mail</strong></td>
</tr>
</tbody>
</table>
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 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.

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 multi-
variate 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 definite-ness. To preserve this feature, many covariance matrix time series models have been proposed based on the Wishart or matrix-F innovation distribution, which generates random positive-definite matrices automatically. 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 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 Stáricá, 2004; Stáricá 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 (Chiriac and Voev, 2011; Corsi, 2009) 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 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, 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...
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, $\text{det}(A)$ 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 $\xi$, $\|\xi\|_k = (E|\xi|^k)^{1/k}$ 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 $B$, $A \otimes^2 = A \otimes A$, $\rightarrow_L$ 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 $y_t \in \mathbb{R}^{n \times n}$ for $t = 1, 2, \ldots, T$ be a positive-definite matrix-valued time series. We consider the following data-generating process (DGP) for $y_t$:

$$y_t = \Sigma_t^{1/2} u_t \Sigma_t^{1/2} \text{ with } \Sigma_t = \Sigma(t/T),$$

(2.1)

where $\Sigma(x)$ is a deterministic bounded positive-definite matrix function with unknown form on the interval $[0,1]$, and $\{u_t\}_{t=1}^T$ is a sequence of strictly stationary positive definite random matrices with $E u_t = I_n$. In model (2.1), both the dynamic process $u_t$ and its parameter matrix $\Sigma_t$ have unspecified forms, allowing the structure of $y_t$ to change abruptly or smoothly. Our idea of using a time-varying function $\Sigma_t$ 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 $y_t$. Moreover, note that the possible nonstationarity of $y_t$ in model (2.1) is caused by the deterministic $\Sigma_t$. This setting is in accordance with the semiparametric model in Section 3, but does not allow $\Sigma_t$ to be nonstationary driven by some unit root-type process. To study the unit root-type nonstationarity of $y_t$, 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 $y_t$, or equivalently, the constancy of $\Sigma_t$. 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 (Chiriac and Voev, 2011; Corsi, 2009). 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 \( \Sigma_0 \) 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 \( y_t \) is unchanged over time. Under \( H_1 \), \( \Sigma_t \) is a time-varying parameter matrix, and allows the structure of \( y_t \) 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 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 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 \( \Sigma_t \), 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 \( \Sigma_t \) is chosen as

\[ \bar{\Sigma} = \frac{1}{T} \sum_{s=1}^{T} y_s, \]

which is obviously efficient and consistent under \( H_0 \), but inconsistent under \( H_1 \). To introduce our second estimator of \( \Sigma_t \), we simply consider the Nadaraya–Watson esti-
mator of $\Sigma(x)$, $x \in [0, 1]$, given by

$$\hat{\Sigma}_T(x) = \frac{\sum_{s=1}^{T} K_h(x - s/T) y_s}{\sum_{s=1}^{T} K_h(x - s/T)}$$

(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 $\frac{1}{T} \sum_{s=1}^{T} K_h(x - s/T) = 1 + O(T^{-1}h^{-1})$ by the integrability of $K(\cdot)$, it is more convenient to consider an alternative estimator of $\Sigma(x)$ given by

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

Given that $E y_t = \Sigma_t$ in model (2.1), we then take our second estimator of $\Sigma_t$ 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 $y_t = y_{-t}$ for $-\lfloor Th \rfloor \leq t \leq -1$, and $y_t = y_{2T-t}$ for $T + 1 \leq t \leq T + \lfloor Th \rfloor$. Using the reflection method, we make the boundary points act similarly to those in the interior region $\lfloor Th \rfloor < t < T - \lfloor Th \rfloor$, and modify $\hat{\Sigma}_t$ as

$$(2.3) \quad \tilde{\Sigma}(x) = \frac{1}{T} \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} K_h(x - s/T) y_s,$$

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

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

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

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

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

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

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

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

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 $T^{-1/2} \sum_{t=1}^{T} \text{vech}(y_t - \Sigma_0)$ (see Andrews (1991); Newey and West (1987)). 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. Statistica Sinica: Preprint doi:10.5705/ss.202021.0029
assumption with the help of instrumental variables. In our model (2.1), we only require \( E(u_t - I_n) = 0 \) instead of the m.d.s. assumption that \( E(u_t - I_n|\mathcal{F}_{t-1}) = 0 \), where \( \mathcal{F}_t = \sigma(y_s, s \leq t) \) 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 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) \( \text{vech}(u_t) \) is a strictly stationary \( \beta \)-mixing process with mixing coefficients \( \beta(j) \) satisfying \( \sum_{j=1}^{\infty} j^2 \beta(j)^{(1+\delta)} < \infty \), for some \( 0 < \delta < 1 \); (ii) \( E\|u_t\|^{4(1+\delta)} < \infty \).

**Assumption 2.3.** (i) \( K : [-1, 1] \to \mathbb{R}^+ \) 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) \( k : [-1, 1] \to \mathbb{R}^+ \) 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 $u_t$. The stationarity condition for $u_t$ 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 $\beta$-mixing condition is widely adopted in the literature. As shown in Proposition 3.1 below, Assumption 2.2(i) holds when $u_t$ follows model (3.1), with $\rho(A_0 + B_0) < 1$. Assumption 2.2(ii) imposes a higher moment condition on $u_t$ so that the limit of $\hat{M}$ under $H_0$ (denoted by $M$) exists and some mixing inequalities can be applied, where

$$M = \Sigma_0^* Z_{\infty} \Sigma_0'$$

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

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

with $z_t = \text{vech}(u_t - I_n)$. Assumption 2.3(i) holds for commonly used kernels, such as the uniform, Epanechnikov, and truncated Gaussian kernels, among 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 $[-1, 1]$. Assumption 2.4(ii) gives a sufficient technical condition on $b_T$ to ensure the consistency of $\hat{M}$ under the null, and was adopted in Hong et al. (2017) and Fu and Hong (2019).

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

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

$$\hat{D} \rightarrow_{\mathcal{L}} N(0, 1) \text{ as } T \rightarrow \infty.$$
Let $c_\alpha$ be the $\alpha$-upper percentile of $N(0,1)$. Because a large value of $\hat{D}$ indicates a rejection of $H_0$, we set the critical region of our test as

\[(2.7) \quad \hat{D} > c_\alpha,\]

which has an asymptotic significance level $\alpha$, by Theorem 2.1.

To compute $\hat{D}$, 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 $t = 1, 2, ..., T$. 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 $y_t$ 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 et al. (2017). In general, to obtain the optimal bandwidth for $\hat{D}$, we may need higher-order Edgeworth expansions for the finite-sample distribution of $\hat{D}$, 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 $Th^{1/2} \hat{S}$ can be estimated precisely. Specifically, we consider the truncated QS kernel for $k$ with $b_T = 1.3221(\pi T)^{1/5}$, where $\pi$ is a function of the unknown spectral density matrix of $\text{vech}(y_t - \Sigma_0)$. 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 \( \pi \) by \( \hat{\pi} \) using preliminary information. For instance, we can take \( \hat{\pi} = \min(\hat{\pi}_1, \hat{\pi}_2) \) with the upper bound \( \hat{\pi}_1 = 1600 \), as suggested by Lima and Xiao (2010), and

\[
(2.8) \quad \hat{\pi}_2 = \sum_{a=1}^{n(n+1)/2} \frac{4(1 + \hat{\phi}_a \hat{\psi}_a)^2(\hat{\phi}_a + \hat{\psi}_a)^2 \hat{\sigma}_a^4}{(1 - \hat{\phi}_a)^6(1 + \hat{\phi}_a^2)} \sqrt{\sum_{a=1}^{n(n+1)/2} \frac{(1 + \hat{\psi}_a)^4 \hat{\sigma}_a^4}{(1 - \hat{\phi}_a)^4}}
\]

(see (6.5) in Andrews (1991)), where \((\hat{\phi}_a, \hat{\psi}_a)\) is the least squares estimator of \((\phi, \psi)\) in the ARMA(1,1) model \( y_{a,t} = \phi y_{a,t-1} + \psi \tau_{a,t-1} + \tau_{a,t}, \) \( \hat{\sigma}_a^2 \) is the estimator of \( \text{Var}(\tau_{a,t}) \), and \( y_{a,t} \) is \( a \)th element of \( \text{vech}(y_t - \Sigma) \). 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

\[ H_{1T}: \Sigma_t = \Sigma_0 + \frac{\Sigma_{1t}}{cT} \text{ for all } t, \]

where \( \Sigma_{1t} = \Sigma_1(t/T) \), the deterministic matrix function \( \Sigma_1(x) \) satisfies the conditions for \( \Sigma(x) \) in Assumption 2.1, and

\[ B_1 = \int_0^1 \text{vech}(\Sigma_1(x))' \text{vech}(\Sigma_1(x))dx - \int_0^1 \text{vech}(\Sigma_1(x))'dx \int_0^1 \text{vech}(\Sigma_1(x))dx > 0. \]

Note that \( \hat{S} \) using the Frobenius norm can be viewed as an \( L_2 \)-type statistic, and its convergence rate in Theorem 2.1 is \( Th^{1/2} \). This implies that the local rate \( cT \) 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 2.2.** 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 \( T \to \infty \),

(i) if \( \frac{T^{1/2}h^{1/4}}{cT} \to 0 \), then \( \hat{D} \to \mathcal{L} N(0,1) \);

(ii) if \( \frac{T^{1/2}h^{1/4}}{cT} \to c \in (0, \infty) \), then \( \hat{D} \to \mathcal{L} N(c^2 B_1, 1) \);

(iii) if \( \frac{T^{1/2}h^{1/4}}{cT} \to \infty \), then \( \hat{D} \to \infty \) in probability;
where $B_{lv} = B_l / \sqrt{N} > 0$, with $N = 2 \text{tr}(M^2) \int \int K(x)K(x + \lambda)dx^2 d\lambda$.

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 $h \to 0$. 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 2.4(ii) is tailored to give a sufficient technical condition on $b_T$ to ensure the consistency of $\hat{M}$ 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 $\hat{D}$, a semiparametric matrix model with an appropriate specification for $u_t$ is needed to model $y_t$. Here, we model $u_t$ by

$$u_t = G_t^{1/2} e_t G_t^{1/2}$$

with $G_t = I_n - A_0 A_0' - B_0 B_0' + A_0 u_{t-1} A_0' + B_0 G_{t-1} B_0'$,

where $G_t \in \mathcal{F}_{t-1}$ has the first-order BEKK specification in Engle and Kroner (1995) with $EG_t = I_n$, and $\{e_t\}_{t=1}^T$ is a sequence of independent and identically distributed (i.i.d.) positive-definite random innovation matrices with $E e_t = I_n$. 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 $u_t \sim (3.1)$ 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, $y_t$ has a nonparametric long-run component $\Sigma_t$ and a parametric short-run component $G_t$, which jointly specify the dynamic structure of $y_t$ with $\mathbb{E}y_t = \Sigma_t$ and $\mathbb{E}(y_t | F_{t-1}) = \Sigma_t^{1/2}G_t\Sigma_t^{1/2}$. 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 dynamic 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 et al. (2021)) and multivariate volatility (see, e.g., Hafner and Linton (2010); Patilea and Raissi (2012); 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 $y_t$ is close to a stationary process $\hat{y}_t(s)$, 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 $A_0 = L_n A_0^{\otimes 2} D_n$ and $B_0 = L_n B_0^{\otimes 2} D_n$. By Theorem 2.1 in Zhou et al. (2021), we have the following proposition, which gives the stationarity and moment properties of $u_t$.

PROPOSITION 3.1. Suppose (i) $e_t$ admits a density $F_{e_t}$, which is absolutely continuous w.r.t. the Lebesgue measure; (ii) the point $I_n$ is in the interior of the support of...
\[ F_e; \text{ (iii) } E \|e_t\| < \infty. \] Then, \( u_t \) is strictly stationary and geometrically ergodic with \( E \|u_t\| < \infty \) if and only if \( \rho(A_0 + B_0) < 1 \).

Note that if \( u_t \) is geometrically ergodic, it is also \( \beta \)-mixing with exponential decay. Hence, if \( \rho(A_0 + B_0) < 1 \), \( y_t \) is \( \beta \)-mixing with exponential decay by Proposition 3.1 and the boundedness of \( \Sigma(x) \). If we do not require \( E \|u_t\| < \infty \), we may use a milder condition than \( \rho(A_0 + B_0) < 1 \) to ensure that \( y_t \) is \( \beta \)-mixing with exponential decay; see, for example, Hafner and Preminger (2009).

Next, we study the moments of \( y_t \) when \( u_t \sim (3.1) \). Let \( u_t = \text{vech}(u_t) \). 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 \( \Phi_0 = I_n \), \( \Phi_i = (A_0 + B_0)^{i-1} A_0 \) and \( r_t = E(u_t | F_{t-1}) - u_t \) is an m.d.s. By using this VMA representation, we can get the cross-covariance of \( u_t \) and \( y_t \); the details are straightforward, and are hence omitted.

**Proposition 3.2.** Suppose \( u_t \) is strictly stationary, with \( E \|u_t\|^2 < \infty \). Then, (i)

\[ E(z_t z_{t-j}') = \sum_{i=0}^{\infty} \Phi_{j+i} E(r_t r_{t-i}') \Phi_i'; \]

(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 \( w_t = \text{vech}(y_t - \Sigma_t) \).

### 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 \( \Gamma_\phi \) 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 \( \phi \).

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 3.1.** \( u_t \) is strictly stationary and ergodic, with \( E \|u_t\|^2 < \infty \).
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 $u_t$, which imply that $u_t$ is $\beta$-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 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

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

Using this notation, 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

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

**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 $\phi$; see Theorem 3.2 and Remark 3 below.

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

$$G_t(\phi) = I_n - AA' - BB' + Au_{t-1}A' + BG_{t-1}(\phi)B', \tag{3.2}$$
for $t = 1, \ldots, T$. Let $\gamma = (\phi', \nu)'$. By assuming that $e_t \sim \nu^{-1}$ Wishart$(\nu, I_n)$ with degrees of freedom $\nu > n$, the negative log-likelihood function (ignoring constants) of $\{y_t\}_{t=1}^T$ is

$$L_T(\gamma) = \nu \sum_{t=1}^T \ell(y_t, \Omega_t(\phi)) + \sum_{t=1}^T c(y_t, \nu) \triangleq \nu \widehat{L}_T(\phi) + \sum_{t=1}^T c(y_t, \nu),$$

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

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

Here, $\nu$ can be viewed as a nuisance parameter, because our main interest is to estimate $\phi$. 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 et al., 2021), the Wishart profiled QMLE of $\phi_0$ does not require any distributional assumption on $e_t$ for its asymptotics.

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

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

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

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

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

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

$$\hat{\gamma} = \arg \min_{\gamma \in \Gamma} \hat{L}_T(\gamma),$$
where $\Gamma = \Gamma_\phi \times \Gamma_\nu$, and $\Gamma_\nu$ is a compact subset of $\{n : \nu > n\}$. Interestingly, we can see that the calculation of $\hat{\phi}$ is irrelevant to $\nu$, 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 $e_t \sim \nu^{-1}\text{Wishart}(\nu, I_n)$, the consistency and asymptotic normality of $\hat{\phi}$ hold for any distribution of $e_t$ with $Ee_t = I_n$ (see Theorem 3.2 below). Hence, we call $\hat{\phi}$ the profiled QMLE of $\phi_0$.

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

**Assumption 3.3.** (i) $\det(A_0) \neq 0$; (ii) $\rho(A + B) < 1$; (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.** $\Gamma_\phi$ is a compact subset of $\mathbb{R}^{2n^2}$, and $\phi_0$ is an interior point of $\Gamma_\phi$.

**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 deriving the asymptotics of $\hat{\phi}$. A restricted $h$ is required to undersmooth the estimation of $\Sigma_t$ so that the impact of its bias on the asymptotics of $\hat{\phi}$ can be neglected; see Hafner and Linton (2010), Jiang et al. (2021), and Remark 4 below for further discussion.

Let $\xi_t = \text{vec}(e_t - I_n)$, $\rho_{t,i} = \text{vec}\left(\frac{\partial G_t}{\partial \phi_i}\right)' \left(G_t^{-1/2}\right)^{\otimes 2}$, $\eta_{t,i} = \text{vec}\left(G_t^{-1} \frac{\partial G_t}{\partial \phi_i}\right)'$, $M^i_j =$

\[\text{Statistica Sinica: Preprint doi:10.5705/ss.202021.0029}\]
\[ E\left( \mathbf{u}_{t-j} \otimes G_t^{-1} \frac{\partial G_t}{\partial \phi} G_t^{-1} \right) \], and

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

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

\[ J_{\phi_0} = E(\rho_t \rho_t') \text{ and } Q_{\phi_0} = N + \Psi + H + H', \]

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

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

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 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) \( \hat{\phi} \to \phi_0 \) in probability as \( T \to \infty; \)

(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,

\[ \sqrt{T}(\hat{\phi} - \phi_0) \to L N(0, V_{\phi_0}) \text{ as } T \to \infty, \]

where \( V_{\phi_0} = J_{\phi_0}^{-1} Q_{\phi_0} J_{\phi_0}^{-1}, \) and \( J_{\phi_0} \) and \( Q_{\phi_0} \) 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 et al. (2021). Therefore, when \( n > 1 \), if there exists cross-sectional dependence between any two elements of \( y_t \), the estimation of \( \Sigma_t \) should affect the asymptotic distribution of \( \hat{\phi} \). Note that when \( n = 1 \), \( \Sigma_t \) represents only the scale of \( y_t \), and has no effect on the dynamic of \( u_t \). 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, the 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, and hence 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 nonparametric estimator \( \hat{\Sigma}_t \) have a negligible effect on the asymptotics of \( \hat{\phi} \). Because of Assumption 3.5, the popular choice of bandwidth \( h = c_T T^{-1/5} \) is excluded. To relax Assumption 3.5, one may use 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 future investigation.

In practice, we can estimate $V_{\phi_0}$ consistently by its sample version. Specifically, let $\hat{G}_t = \hat{G}_t(\hat{\phi})$, $\frac{\partial \hat{G}_t}{\partial \phi_i} = \hat{\varepsilon}_i = \hat{G}_t^{-1/2}\hat{u}_t\hat{G}_t^{-1/2}$ and define $\hat{\eta}_t$, $\hat{\rho}_t$, $\hat{\xi}_t$ accordingly. Based on these quantities, we compute the sample versions of $J_{\phi_0}$ and $M_{ij}$ by

$$\hat{J} = \frac{1}{T} \sum_{t=1}^{T} \hat{\rho}_t \hat{\rho}'_t,$$

$$\hat{M}_{ij} = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t \otimes \hat{G}_t^{-1/2} \frac{\partial \hat{G}_t}{\partial \phi_i} \hat{G}_t^{-1/2}$$

and $\hat{M}_j = \frac{1}{T} \sum_{t=j+1}^{T} \hat{u}_t \otimes \hat{G}_t^{-1/2} \frac{\partial \hat{G}_t}{\partial \phi_i} \hat{G}_t^{-1/2}$, respectively. To compute the sample version of $Q_{\phi_0}$, we first let

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

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

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

and $b_T$ is defined in Assumption 2.4(ii). Based on $\hat{\Upsilon}(t/T)$, we next compute the sample versions of $N$, $\Psi$, and $H$ by

$$\text{vec}(\hat{N}) = \frac{1}{T} \sum_{t=1}^{T} [\hat{\rho}_t \otimes 2] \text{vec}(\text{Var}(\xi_t)),$$

$$\text{vec}(\hat{\Psi}) = \left\{ \frac{1}{T} \sum_{t=1}^{T} [\hat{\Upsilon}(t/T) \otimes \hat{\Upsilon}(t/T)] \right\} [\hat{\Sigma}_{1/4} \otimes I_{2n^2}] \left\{ \frac{1}{T} \sum_{t=1}^{T} [(\hat{G}_t^{1/2}) \otimes 2] \otimes \hat{\rho}_t \right\} \text{vec}(\text{Var}(\xi_t)),$$

$$\text{vec}(\hat{H}) = \left\{ \frac{1}{T} \sum_{t=1}^{T} [\hat{\Upsilon}(t/T) \otimes I_{2n^2}] \right\} [\hat{\Sigma}_{1/4}] \frac{1}{T} \sum_{t=1}^{T} \left\{ [(\hat{G}_t^{1/2}) \otimes 4] \right\} \text{vec}(\text{Var}(\xi_t)),$$

respectively, where $\text{Var}(\xi_t)$ is the sample variance of $\{\hat{\xi}_t\}$. Further, we compute the sample version of $Q_{\phi_0}$ by $\hat{Q}$, where $\hat{Q} = \hat{\Psi} + \hat{H} + \hat{H}' + \hat{N}$. Finally, we obtain the sample version of $V_{\phi_0}$ as $\hat{V} := \hat{J}^{-1}\hat{Q}\hat{J}^{-1}$.

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

$$(3.9) \quad \hat{W} = T (\Gamma_{\phi} - \iota) (\Gamma\hat{V}\Gamma')^{-1} (\Gamma_{\phi} - \iota)$$
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 $\hat{W}$ is larger than the upper-tailed critical value of $\chi^2_s$.

It is also of interest to estimate the nuisance parameter $\nu$ by $\hat{\nu}$ in (3.6), if we have prior information that $e_t \sim \nu_0^{-1}$ 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

$$\hat{\nu} = \arg \min_{\nu \in \Gamma \nu} \hat{L}_T(\hat{\phi}, \nu) .$$

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

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

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

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

By some simple calculations, we can see that when $e_t \sim \nu_0^{-1}$ Wishart($I_n, \nu_0$), the Cramér–Rao lower bound with respect to $\nu_0$ is $2J_{\nu_0}^{-1}$. Interestingly, Theorem 3.3(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 1000 replications of sample
size $T = 500, 1000, 1500, 3000$ from the following two DGPs:

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

**DGP 2:** The TVB model with $A_0 = \begin{pmatrix} 0.5 & 0 \\ 0.2 & 0.4 \end{pmatrix}$ 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) \[ \text{[No change]} \quad \Sigma(x) = 2I_2, \]

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

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

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

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

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

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

[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 $F(\nu_0^\dagger, \frac{\nu_20 - n - 1}{\nu_10} I_n)$ 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 $\hat{D}$ according to the statements in Subsection 2.3. Specifically, we consider three choices of kernel function $K(x)$:

- **Uniform (U)**: $K(x) = \frac{1}{2} 1(|x| \leq 1),$
- **Epanechnikov (E)**: $K(x) = \frac{3}{4} (1 - x^2) 1(|x| \leq 1),$
- **Truncated Normal (TN)**: $K(x) = \frac{\exp(-x^2) - \exp(-1/2)}{\text{erf}(1/\sqrt{2})\sqrt{2\pi} - 2 \exp(-1/2)} 1(|x| \leq 1),$

and take the bandwidth $h = \left(1/\sqrt{12}\right) T^{-1/5}$, where $\text{erf}(\cdot)$ is the error function. When $\hat{D}$ falls into the critical region in (2.7), we reject $H_0$ at level $\alpha$.

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 $l_B^T$ such that $l_B^T \to \infty$ and $l_B^T/T \to 0$ as $T \to \infty$;
2. Divide the original data into $T - l_B^T + 1$ blocks $\{\Xi_t\}_{t=1}^{T-l_B^T+1}$, where $\Xi_t = \{y_t, ..., y_{t+l_B^T-1}\}$;   
3. Resample $\{\Xi_t\}_{t=1}^{T-l_B^T+1}$ with replacement to form the bootstrapped data $\{\Xi_t^*\}_{t=1}^{T-l_B^T+1}$;
4. Calculate $\hat{D}^*$ based on the bootstrapped data in step 3;
5. Repeat steps 3–4 $J$ times to get the bootstrapped test statistics $\{\hat{D}_{j}^*\}_{j=1}^{J}$, calculate the empirical 100$(1 - \alpha)$ sample percentile of $\{\hat{D}_{j}^*\}_{j=1}^{J}$ (denoted by $\hat{D}^*_\alpha$), and set the bootstrapped critical region as

\begin{equation}
\hat{D} > \hat{D}^*_\alpha; 
\end{equation}

6. Reject $H_0$ at level $\alpha$ when the value of $\hat{D}$ falls into the critical region in (4.6).
To implement this moving block method, we follow Politis and White (2004) to take
\[ l_T^B = \left[ \frac{3}{2} c_B T \right]^{1/3}, \]
where \( c_B = \text{tr} \left( \sum_{j=-b_T}^{b_T} k\left( \frac{j}{b_T} \right) |j| \hat{\Gamma}_{v,j} \right) / \text{tr} \left( \hat{M} \right), \]
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 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 \( \hat{D} \), where \( \hat{D}^{AC} \) and \( \hat{D}^{BC} \) 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, \( \hat{D}^{AC} \) is undersized, but the size of \( \hat{D}^{BC} \) is more accurate; (ii) both \( \hat{D}^{AC} \) and \( \hat{D}^{BC} \) have satisfactory size performance, regardless of the choice of \( K(x) \) and the distribution of \( e_t \), when the sample size \( T \) increases, and \( \hat{D}^{BC} \) has slightly better size performance than \( \hat{D}^{AC} \), especially at level 10%; and (iii) the size of \( \hat{D}^{BC} \) 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 \( \hat{D} \) under the alternative settings in (4.2)–(4.5). As before, we consider both \( \hat{D}^{AC} \) and \( \hat{D}^{BC} \). Their results at level 5% are reported in Fig 1 for the data generated from DGP 1, where \( T = 1500 \) 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 size. Considering the trade-off between size accuracy and power enhancement, we recommend considering the results of both $\hat{D}^{AC}$ and $\hat{D}^{BC}$ in practice.

4.2. Study on the estimation. In this subsection, we assess the finite-sample performance of the profiled QMLE $\hat{\phi}$. 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 $r = 10$. To compute $\hat{\phi}$ 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.

\begin{table}[h]
\centering
\caption{Sizes of $\hat{D}^{AC}$ and $\hat{D}^{BC}$ at level $\alpha = 5\%$ and $10\%$.}
\begin{tabular}{llllllllllllll}
\hline
 & \multicolumn{6}{c}{$e_t \sim \text{Wishart}$} & \multicolumn{6}{c}{$e_t \sim \text{matrix-F}$} \\
 & $T$ & $K(x)$ & 5\% & 10\% & 5\% & 10\% & 5\% & 10\% & 5\% & 10\% & 5\% & 10\% & 5\% & 10\% \\
\hline
DGP 1 & & & & & & & & & & & & & & \\
500 & U & 0.025 & 0.045 & 0.026 & 0.038 & 0.024 & 0.038 & 0.023 & 0.037 & 0.019 & 0.038 & 0.019 & 0.037 \\
 & E & 0.020 & 0.042 & 0.045 & 0.090 & 0.024 & 0.038 & 0.036 & 0.065 & 0.019 & 0.038 & 0.019 & 0.037 \\
 & TN & 0.019 & 0.043 & 0.044 & 0.088 & 0.023 & 0.038 & 0.037 & 0.065 & 0.022 & 0.043 & 0.036 & 0.077 \\
1000 & U & 0.040 & 0.073 & 0.042 & 0.092 & 0.041 & 0.070 & 0.051 & 0.105 & 0.042 & 0.083 & 0.042 & 0.087 \\
 & E & 0.042 & 0.060 & 0.051 & 0.092 & 0.022 & 0.043 & 0.036 & 0.077 & 0.022 & 0.040 & 0.038 & 0.079 \\
 & TN & 0.042 & 0.059 & 0.050 & 0.093 & 0.022 & 0.040 & 0.038 & 0.079 & 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 & 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 & 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 & 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 & 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 & 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 & 0.046 & 0.079 & 0.045 & 0.088 \\
\hline
DGP 2 & & & & & & & & & & & & & & \\
500 & U & 0.036 & 0.067 & 0.054 & 0.106 & 0.018 & 0.036 & 0.047 & 0.080 & 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 & 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 & 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 & 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 & 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 & 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 & 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 & 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 & 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 & 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 & 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 & 0.046 & 0.077 & 0.045 & 0.083 \\
\hline
\end{tabular}
\end{table}
Fig 1. Power across $r$ for $\hat{D}^{AC}$ (triangle “$\triangle$” marker) and $\hat{D}^{BC}$ (cross “×” marker) when $T = 1500$ (solid line) and $T = 3000$ (dotted line). The horizontal dashed line corresponds to the level $\alpha = 5\%$.

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 3.5.

Tables 2 and 3 report the sample bias, sample empirical standard deviation (ESD), and average asymptotic standard deviation (ASD) of $\hat{\phi}$ 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 $\hat{\phi}$ are small in all considered cases; (ii) regardless of the specification of $\Sigma(x)$ and the distribution of $e_t$, 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 $e_t \sim$ Wishart are smaller than the corresponding values of ESD (or ASD) in the case of $e_t \sim$ matrix-F, and this is consistent with the fact that $\hat{\phi}$ is an efficient estimator when $e_t$ follows a Wishart distribution. Note that when $e_t \sim$ 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.

### Table 2

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

<table>
<thead>
<tr>
<th>$e_t$</th>
<th>$T$</th>
<th>$A_{11}$</th>
<th>$A_{21}$</th>
<th>$A_{12}$</th>
<th>$A_{22}$</th>
<th>$B_{11}$</th>
<th>$B_{21}$</th>
<th>$B_{12}$</th>
<th>$B_{22}$</th>
<th>$\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wishart</td>
<td>1500</td>
<td>Bias</td>
<td>-0.0055</td>
<td>0.0030</td>
<td>-0.0011</td>
<td>-0.0070</td>
<td>-0.0093</td>
<td>-0.0138</td>
<td>0.0102</td>
<td>-0.0090</td>
</tr>
<tr>
<td>ESD</td>
<td>0.0172</td>
<td>0.0169</td>
<td>0.0185</td>
<td>0.0189</td>
<td>0.0259</td>
<td>0.0248</td>
<td>0.0285</td>
<td>0.0299</td>
<td>0.4229</td>
<td></td>
</tr>
<tr>
<td>ASD</td>
<td>0.0156</td>
<td>0.0176</td>
<td>0.0180</td>
<td>0.0187</td>
<td>0.0255</td>
<td>0.0245</td>
<td>0.0302</td>
<td>0.0279</td>
<td>0.4089</td>
<td></td>
</tr>
<tr>
<td>3000</td>
<td>Bias</td>
<td>-0.0021</td>
<td>0.0017</td>
<td>-0.0006</td>
<td>-0.0030</td>
<td>-0.0056</td>
<td>0.0080</td>
<td>0.0051</td>
<td>-0.0047</td>
<td>0.0504</td>
</tr>
<tr>
<td>ESD</td>
<td>0.0118</td>
<td>0.0118</td>
<td>0.0131</td>
<td>0.0133</td>
<td>0.0176</td>
<td>0.0166</td>
<td>0.0203</td>
<td>0.0199</td>
<td>0.2903</td>
<td></td>
</tr>
<tr>
<td>ASD</td>
<td>0.0110</td>
<td>0.0122</td>
<td>0.0126</td>
<td>0.0131</td>
<td>0.0176</td>
<td>0.0166</td>
<td>0.0206</td>
<td>0.0191</td>
<td>0.2886</td>
<td></td>
</tr>
<tr>
<td>matrix-F</td>
<td>1500</td>
<td>Bias</td>
<td>-0.0319</td>
<td>0.0009</td>
<td>0.0019</td>
<td>-0.0245</td>
<td>0.0096</td>
<td>-0.0102</td>
<td>0.0098</td>
<td>-0.0009</td>
</tr>
<tr>
<td>ESD</td>
<td>0.0315</td>
<td>0.0306</td>
<td>0.0306</td>
<td>0.0337</td>
<td>0.0436</td>
<td>0.0394</td>
<td>0.0412</td>
<td>0.0413</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASD</td>
<td>0.0296</td>
<td>0.0330</td>
<td>0.0324</td>
<td>0.0409</td>
<td>0.0425</td>
<td>0.0352</td>
<td>0.0384</td>
<td>0.0363</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3000</td>
<td>Bias</td>
<td>-0.0028</td>
<td>0.0012</td>
<td>0.0029</td>
<td>-0.0169</td>
<td>-0.0119</td>
<td>-0.0062</td>
<td>0.0028</td>
<td>0.0048</td>
<td></td>
</tr>
<tr>
<td>ESD</td>
<td>0.0229</td>
<td>0.0224</td>
<td>0.0224</td>
<td>0.0299</td>
<td>0.0315</td>
<td>0.0269</td>
<td>0.0207</td>
<td>0.0322</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASD</td>
<td>0.0250</td>
<td>0.0236</td>
<td>0.0168</td>
<td>0.0303</td>
<td>0.0309</td>
<td>0.0242</td>
<td>0.0206</td>
<td>0.0248</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel A: $\Sigma(x) \sim (4.2)$ with $r = 10$

| Wishart | 1500 | Bias | -0.0400 | 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 | 0.0114 | 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 | |

Panel B: $\Sigma(x) \sim (4.5)$ with $r = 10$

### 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 et al. (2016). This data set ranges from January 2006 to December 2011, with 1474 observations in total. Two flash crashes, on May 6, 2010, and August 9, 2011, are replaced by the average of their nearest five preceding and following matrices, respectively. Denote this data set by $\{y_{t}\}_{t=1}^{1474}$. First, we plot $\{y_{rs,t}\}_{t=1}^{1474}$ and the sample ACFs of $y_{rs,t}$ in Fig 2, where $y_{rs,t}$ is the $(r,s)$th element of $y_t$. This figure shows a peak in all $y_{rs,t}$ around late 2008, and the sample ACFs of each $y_{rs,t}$ decay slowly.

Next, we apply our generalized Hausman test $\hat{D}$ to detect whether there is any
structural change in $y_t$. 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 $y_t$ is changing over time.

Third, we fit $y_t$ 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 $\hat{\Sigma}_{rs,t}$ (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 $\Sigma_t$ 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 \{$\hat{e}_{rs,t}$\} (denoted by $\rho_{rs}(j)$ for $j = 1, \ldots, 100$) in the three bottom panels of Fig 3, where \(\hat{e}_{rs,t}\) is the \((r, s)\)th element of \(\hat{e}_t\). 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 $e_{12,t}$ and $e_{22,t}$ 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:

\[ y_t = G_t^{1/2}e_tG_t^{1/2}, \]

\[ G_t = \Omega + A(d)y_{t-1,d}A'(d) + A(w)y_{t-1,w}A'(w) + A(m)y_{t-1,m}A'(m), \]
where $y_{t-1,d} = y_{t-1}$, $y_{t-1,w} = (1/5) \sum_{i=1}^{5} y_{t-i}$, and $y_{t-1,m} = (1/22) \sum_{i=1}^{22} y_{t-i}$ 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 $e_t \sim \nu^{-1} \text{Wishart}(\nu, I_n)$. 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 \( \{y_t\}_{t=T_0-999}^{T_0} \) to make an \( l \)-step-ahead forecast $\hat{y}_{T_0+l|T_0}$ for the out-of-sample data point $y_{T_0+l}$, where $T_0 = 1000, \ldots, 1474 - l$. For the TVB model, $\hat{y}_{T_0+l|T_0}$ is calculated as $\hat{y}_{T_0+l|T_0} = \hat{\Sigma}_{T_0}^{1/2} \hat{u}_{T_0+l|T_0} \hat{\Sigma}_{T_0}^{1/2}$, where $\hat{\Sigma}_{T_0}$ is the estimate of $\Sigma_{T_0}$, and $\hat{u}_{T_0+l|T_0}$ is the forecast of $u_{T_0+l}$. For the HAR model, $\hat{y}_{T_0+l|T_0}$ is calculated as $\hat{y}_{T_0+l|T_0} = \hat{G}_{T_0+l|T_0}$, where $\hat{G}_{T_0+l|T_0}$ is the forecast of $G_{T_0+l}$. Table 4 reports the average of forecast errors $\{\hat{y}_{T_0+l|T_0} - y_{T_0+l}\}$ 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 $y_t$ into account.

**Table 4**

<table>
<thead>
<tr>
<th></th>
<th>( l = 1 )</th>
<th>( l = 5 )</th>
<th>( l = 10 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frobenius</td>
<td>Spectral</td>
<td>Frobenius</td>
</tr>
<tr>
<td>TVB model</td>
<td>0.9137</td>
<td>0.8873</td>
<td>1.1978</td>
</tr>
<tr>
<td>HAR model</td>
<td>0.9183</td>
<td>0.8900</td>
<td>1.2235</td>
</tr>
</tbody>
</table>
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
et al. (2019)) could be another promising direction for future research.

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.2019207L01009). 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 Supplementary Material contains Appendices A, B, C, and D, where Appendix A gives proofs of Theorems 2.1–2.2, Appendix B offers proofs of Theorems 3.1–3.3, Appendix C lists some basic derivatives results, and Appendix D provides some numerical evidence of spurious long-memory phenomena caused by structural changes.

References.


Fudan University  
Department of Statistics and Data Science  
Shanghai  
China  
E-MAIL: jiangfy@fudan.edu.cn

Tsinghua University  
Department of Industrial Engineering and Center for Statistical Science  
Beijing  
China  
E-MAIL: malidong@tsinghua.edu.cn
STRUCTURAL CHANGE IN COVARIANCE MATRIX TIME SERIES

Education University of Hong Kong
Department of Mathematics and Information Technology
Hong Kong
China
E-mail: waikeungli@eduhk.hk

University of Hong Kong
Department of Statistics & Actuarial Science
Hong Kong
China
E-mail: mazhuke@hku.hk