Statistica Sinica **32** (2022), 755-786
doi:https://doi.org/10.5705/ss.202019.0424

TIME SERIES MODELS FOR REALIZED COVARIANCE MATRICES BASED ON THE MATRIX-F DISTRIBUTION

Jiayuan Zhou[1], Feiyu Jiang[2], Ke Zhu[3] and Wai Keung Li[3,4]

[1] *University of Florida,* [2] *Fudan University,*
[3] *The University of Hong Kong and* [4] *The Education University of Hong Kong*

Abstract: We propose a new Conditional BEKK matrix-F (CBF) model for time-varying realized covariance (RCOV) matrices. This CBF model is capable of capturing a heavy-tailed RCOV, which is an important stylized fact, but is not handled adequately by Wishart-based models. To further mimic the long-memory feature of an RCOV, we introduce a special CBF model with a conditional heterogeneous autoregressive structure. Moreover, we provide a systematic study of the probabilistic properties and statistical inferences of the CBF model, including exploring its stationarity, establishing the asymptotics of its maximum likelihood estimator, and giving new inner-product-based tests for model checking. In order to handle a large-dimensional RCOV matrix, we construct two reduced CBF models: the variance-target CBF model (for a moderate but fixed-dimensional RCOV matrix), and the factor CBF model (for a high-dimensional RCOV matrix). For both reduced models, the asymptotic theory of the estimated parameters is derived. The importance of our methodology is illustrated by means of simulations and two real examples.

Key words and phrases: Factor model, heavy-tailed innovation, long memory, matrix-F distribution, matrix time series model, model checking, realized covariance matrix, variance target.

1. Introduction

Modeling the multivariate volatility of many asset returns is crucial for asset pricing, portfolio selection, and risk management. Since the seminal work of Barndorff-Nielsen and Shephard (2002, 2004) and Andersen et al. (2003), the realized covariance (RCOV) matrix, estimated from intraday high-frequency return data, has been recognized as better than the daily squared returns as an estimator for daily volatility. Consequently, attention has increased on the modeling and forecasting of these RCOVs; see, for example, McAleer and Medeiros (2008), Hansen, Huang and Shek (2012), Noureldin, Shephard and Sheppard (2012) and Bollerslev, Patton and Quaedvlieg (2016), among many others.

Corresponding author: Feiyu Jiang, Department of Statistics and Data Science, Fudan University, Shanghai, China. E-mail: jiangfy@fudan.edu.cn.

Existing models for RCOV matrices can be roughly categorized into two types: transformation-based models and likelihood-based models. Models in the first category capture the dynamics of the RCOV matrices in an indirect way via transformation. Bauer and Vorkink (2011) used a factor model for the vectorization of the log transformation of an RCOV matrix; Chiriac and Voev (2011) applied a vector autoregressive fractionally integrated moving average process to model the Cholesky decomposition of an RCOV matrix; and Callot, Kock and Medeiros (2017) transformed the RCOV matrix into a large vector using the *vech* operator, and then fitted this transformed vector using a vector autoregressive model. In the first two models, the dimension of the RCOV matrix has to be moderate (e.g., less than six) for a feasible manipulation. In the third model, the dimension of the RCOV matrix is allowed to be thirty in applications with the help of the LASSO method.

Models in the second category deal with RCOV matrices directly by assuming that the innovation driving the RCOV time series has a specific matrix distribution in order to generate random positive-definite matrices automatically, without imposing additional constraints. This important feature results in positive-definite estimated RCOV matrices. Unlike scalar or vector distributions, so far, few matrix distributions have been found to have explicit forms. The primary choice for the innovation distribution is Wishart, leading to the Wishart autoregressive (WAR) model in Gouriéroux, Jasiak and Sufana (2009), the conditional autoregressive Wishart (CAW) model in Golosnoy, Gribisch and Liesenfeld (2012), the mixture Wishart model in Jin and Maheu (2013, 2016), and the generalized CAW model in Yu, Li and Ng (2017), to name a few. The other choice for the innovation distribution is matrix-F, which was recently adopted by Opschoor et al. (2018). In general, the matrix-F distribution is a generalization of the usual F distribution whereas the Wishart distribution is a generalization of the χ^2 distribution (see, e.g., Konno (1991) and Opschoor et al. (2018)). Therefore, the matrix-F distribution could be more appropriate than the Wishart distribution in terms of capturing the heavy-tailed innovation, which is an important stylized fact in many applications (see, e.g., Bollerslev (1987), Fan, Qi and Xiu (2014), Zhu and Li (2015), and Oh and Patton (2017)). These likelihood models have at least three advantages over transformation-based models. First, likelihood-based models preserve useful and important matrix structural information, which makes them more interpretable than transformation-based models. Second, the number of estimated parameters in transformation-based models is $O(n^4)$, whereas that of likelihood-based models is $O(n^2)$, where n is the dimension of the RCOV matrix. When n is large, likelihood-based models can be more convenient and less daunting in terms of computation. Third, likelihood-based models use the likelihood function of the RCOV matrices, so that their statistical inference methods are easily provided.

This paper contributes to the literature in three ways. First, we propose a new Conditional BEKK matrix-F (CBF) model with which to study time-varying RCOV matrices. Our CBF model has matrix-F distributed innovations with two degrees of freedom parameters, ν_1 and ν_2. When $\nu_2 \to \infty$, our CBF model reduces to the CAW model (Golosnoy, Gribisch and Liesenfeld (2012)), which has Wishart distributed innovations. Hence, ν_2 is designed to capture the heavy-tailedness of the RCOV. Because an RCOV has been shown to have a long-memory feature, we further introduce a special CBF model that has a similar conditional heterogeneous autoregressive (HAR) structure, as in Corsi (2009). This special model is referred to as the CBF-HAR model. Although the CBF-HAR model is not formally a long-memory model, it gives rise to persistence in the RCOV time series. Two real examples demonstrate that our CBF model (especially the CBF-HAR model) can exhibit significantly better forecasting performance than the corresponding CAW model. Hence, a simple incorporation of ν_2 to capture the heavy-tailed RCOV is necessary from a practical viewpoint.

Second, we provide a systematic statistical inference procedure for the CBF model. Specifically, we explore its stationarity conditions, establish the strong consistency and asymptotic normality of its maximum likelihood estimator (MLE), and investigate some new inner-product-based tests for model diagnostic checking. Moreover, the performance of our methodology is assessed using simulation studies. Compared with those for existing BEKK-type multivariate time series models, our proofs for the CBF model are much more involved, because the CBF model is tailored for matrix time series. In particular, our inner-product-based tests seem to be the first diagnostic checking tool for matrix time series models, and can be extended easily to other models.

Third, we construct two reduced CBF models, the variance targeted (VT) CBF (VT-CBF) model and the factor CBF (F-CBF) model, to handle moderately large and high-dimensional RCOV matrices, respectively. For both reduced models, we derive the asymptotic theory of the estimated parameters. The dimension of the RCOV matrix is allowed to be a moderate, but fixed number in the VT-CBF model, while it is allowed to grow with the sample size T and the intraday sample size in the F-CBF model. This makes the prediction of large-dimensional RCOV matrices feasible in many cases. The importance of both reduced models is illustrated by means of two real applications.

The remainder of the paper is organized as follows. Section 2 introduces the

CBF model and studies its probabilistic properties. Section 3 investigates the asymptotics of the MLE. Section 4 presents inner-product-based tests to check the model adequacy. Two reduced CBF models and their related asymptotic theories are provided in Section 5. Some simulation studies are carried out in Section 6. Applications are given in Section 7. Section 8 concludes this paper. The proofs of all theorems are relegated to the Supplementary Material.

The following notation is used throughout the paper. I_n is the identity matrix of order n, and $\otimes$ represents the Kronecker product. For an $n \times n$ matrix A, $tr(A)$ is its trace, A' is its transpose, $|A|$ is its determinant, $\rho(A)$ is its biggest eigenvalue, $\|A\| = \sqrt{tr(A'A)}$ is its Euclidean (or Frobenius) norm, $\|A\|_{spec} = \sqrt{\rho(A'A)}$ is its spectral norm, $vec(A)$ is a vector obtained by stacking all the columns of A, $vech(A)$ is a vector obtained by stacking all columns of the lower triangular part of A, and $A^{\otimes 2} = A \otimes A$.

2. Model and Properties

2.1. Model specification

Let Y_t^* be the integrated volatility matrix of n asset returns X_t at time $t = 1, \ldots, T$. Since the seminal work of Barndorff-Nielsen and Shephard (2002, 2004) and Andersen et al. (2003), the $n \times n$ positive-definite RCOV matrix Y_t, calculated from the high-frequency return data of X_t, has been widely applied to estimate Y_t^* in the literature; see, for example, Barndorff-Nielsen, Hansen and Lunde (2011), Lunde, Shephard and Sheppard (2016), Aït-Sahalia and Xiu (2017), Kim et al. (2018), and the references therein. Moreover, Y_t is often viewed as a precise estimate for the conditional variances and covariances of these n low-frequency asset returns X_t; hence, how to predict Y_t using some dynamic models is important in practice. Motivated by this, we propose a new dynamic model for Y_t.

Let $\mathcal{G}_t = \sigma(Y_s; s \le t)$ be a filtration up to time t. We assume that

$$Y_t = \Sigma_t^{1/2} \Delta_t \Sigma_t^{1/2}, \tag{2.1}$$

where $\{\Delta_t\}_{t=1}^T$ is a sequence of independent and identically distributed (i.i.d.) $n \times n$ positive-definite random innovation matrices with $E(\Delta_t | \mathcal{G}_{t-1}) = I_n$, each Δ_t follows the matrix-F distribution $F(\nu, ((\nu_2 - n - 1)/\nu_1) I_n)$, and the density of $F(\nu, \Sigma)$ is

$$f(x; \nu, \Sigma) = \Lambda(\nu) \times \frac{|\Sigma|^{-\nu_1/2} \, |x|^{(\nu_1 - n - 1)/2}}{|I_n + \Sigma^{-1} x|^{(\nu_1 + \nu_2)/2}}, \quad \text{for } x \in \mathcal{R}^{n \times n}, \tag{2.2}$$

where $\nu = (\nu_1, \nu_2)'$ with degrees of freedom $\nu_1 > n+1$ and $\nu_2 > n+1$, Σ is an $n \times n$ positive-definite matrix, and

$$\Lambda(\nu) = \frac{\Gamma_n((\nu_1+\nu_2)/2)}{\Gamma_n(\nu_1/2)\Gamma(\nu_2/2)} \quad \text{with} \quad \Gamma_n(x) = \pi^{n(n-1)/4} \prod_{i=1}^{n} \Gamma\left(x + \frac{1-i}{2}\right).$$

Moreover, $\Sigma_t^{1/2} \in \mathcal{G}_{t-1}$ is the square root of the $n \times n$ positive-definite matrix Σ_t, which has a BEKK-type dynamic structure (see Engle and Kroner (1995)):

$$\Sigma_t = \Omega + \sum_{i=1}^{P} \sum_{k=1}^{K} A_{ki} Y_{t-i} A'_{ki} + \sum_{j=1}^{Q} \sum_{k=1}^{K} B_{kj} \Sigma_{t-j} B'_{kj}, \tag{2.3}$$

where Ω, A_{ki}, and B_{kj} are all $n \times n$ real matrices, the integers P, Q, and K are known as the orders of the model, and Ω and the initial states $\Sigma_0, \Sigma_{-1}, \ldots, \Sigma_{-Q+1}$ are all positive definite. Under model (2.1),

$$Y_t | \mathcal{G}_{t-1} \sim F\left(\nu, \frac{\nu_2 - n - 1}{\nu_1} \Sigma_t\right), \tag{2.4}$$

with $E(Y_t | \mathcal{G}_{t-1}) = \Sigma_t$; that is, the conditional distribution of Y_t is matrix-F with a BEKK-type mean structure. As such, we call model (2.1) the Conditional BEKK matrix-F (CBF) model.

The CBF model is related to the CAW model of Golosnoy, Gribisch and Liesenfeld (2012), in which Δ_t follows the Wishart distribution. To see this clearly, we follow Konno (1991) and Leung and Lo (1996) by rewriting Y_t in model (2.1) as

$$Y_t = \left(\frac{\nu_2 - n - 1}{\nu_1}\right) \Sigma_t^{1/2} L_t^{1/2} R_t^{-1} L_t^{1/2} \Sigma_t^{1/2}, \tag{2.5}$$

where $L_t \sim \text{Wishart}(\nu_1, I_n)$ and $R_t \sim \text{Wishart}(\nu_2, I_n)$ are independent. Because $\lim_{\nu_2 \to \infty} \nu_2^{-1} R_t = I_n$ in probability, the identity (2.5) implies that when $\nu_2 \to \infty$, $Y_t | \mathcal{G}_{t-1} \sim \text{Wishart}(\nu_1,\ \nu_1^{-1} \Sigma_t)$, which is exactly the CAW model. Therefore, compared with the CAW model, the degrees of freedom ν_2 in the CBF model accommodate the heavy-tailed RCOV, meaning that each $Y_{t,ij}$ from Y_t satisfying (2.4) could have a heavier tail than that from Y_t satisfying $Y_t | \mathcal{G}_{t-1} \sim \text{Wishart}(\nu_1, \nu_1^{-1} \Sigma_t)$ (see, e.g., Opschoor et al. (2018) for more discussion and examples). Clearly, the identity (2.5) also guarantees Y_t to be symmetric and positive definite, and can be used to generate Y_t by using Wishart random variables.

In addition to the heavy-tailedness, long memory is another well-documented feature of the RCOV, and has been taken into account by many RCOV models, including the HAR model of Corsi (2009) as a benchmark. Although the HAR model does not formally belong to the class of long-memory models, it is able to reproduce the persistence of RCOVs observed in empirical data. Inspired by the HAR model, we consider a special CBF model with the following specification for Σ_t:

$$\Sigma_t = \Omega + A_{(d)}Y_{t-1,d}A'_{(d)} + A_{(w)}Y_{t-1,w}A'_{(w)} + A_{(m)}Y_{t-1,m}A'_{(m)}, \tag{2.6}$$

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 the RCOV matrices. In this case, we label model (2.1) as the CBF-HAR model, because we put "HAR dynamics" on Σ_t. Clearly, the CBF-HAR model is simply a constrained CBF model with $P = 22$, $K = 3$, and $Q = 0$. Figure 1 plots the sample autocorrelation functions (ACFs) up to lag 100 for simulated data from the CBF-HAR model with $\nu = (20, 10)$ and

$$\Omega = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.2 & 0.5 & 0.25 \\ 0.3 & 0.25 & 0.5 \end{pmatrix}, \qquad A_{(d)} = \begin{pmatrix} 0.7 & 0 & 0 \\ 0 & 0.65 & 0 \\ 0 & 0 & 0.75 \end{pmatrix},$$

$$A_{(w)} = \begin{pmatrix} 0.6 & 0 & 0 \\ 0 & 0.6 & 0 \\ 0 & 0 & 0.55 \end{pmatrix}, \qquad A_{(m)} = \begin{pmatrix} 0.4 & 0 & 0 \\ 0 & 0.45 & 0 \\ 0 & 0 & 0.4 \end{pmatrix}.$$

The figure shows that all entries of Y_t exhibit the long-memory feature, as expected.

Note that when $K = 1$, the sufficient identifiability conditions of model (2.3) are that the main diagonal elements of Ω and the first diagonal element of each A_{1i}, B_{1j} are positive; when $K > 1$, some sufficient identifiability conditions of model (2.3) can be found in Engle and Kroner (1995). For simplicity, we assume subsequently that model (2.3) is identifiable.

Of course, the BEKK specification in model (2.3) is not the only way to describe the dynamics of Σ_t. Multivariate ARCH-type models, such as the VEC model of Bollerslev, Engle and Wooldridge (1988), component model of Engle and Lee (1999), and dynamic conditional correlation model of Engle (2002), among many others, can also be adopted to model Σ_t. Using these models together with the matrix-F distribution to fit and predict the RCOV matrices could be a promising direction for future study.

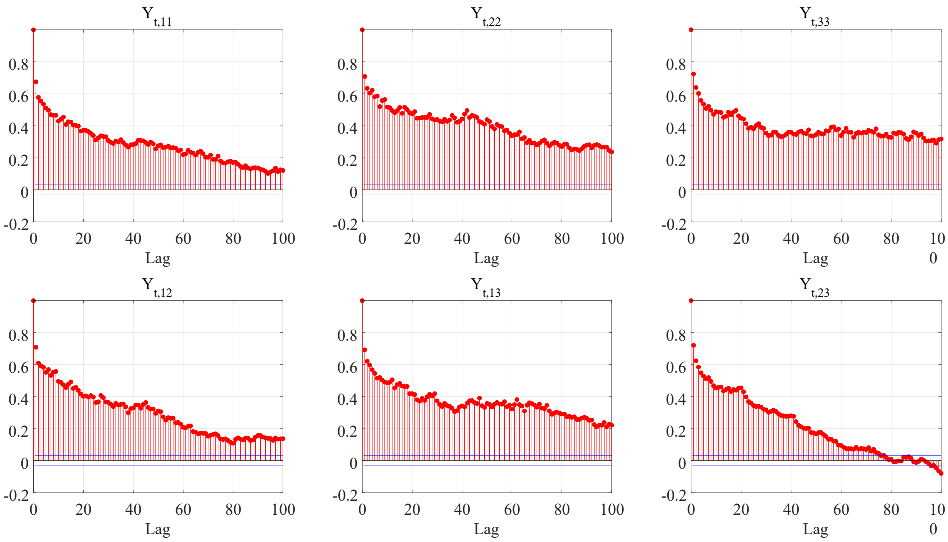

Figure 1. Sample ACFs for simulated data from a 3×3 CBF-HAR model

2.2. Stationarity

Stationarity is an important issue for most RCOV models, but so far has been rarely studied. Denote $M = max(P, Q)$. For $i = 1, 2, \ldots, M$, let

$$A_i^* = \sum_{k=1}^{K} A_{ik}^{\otimes 2} \quad \text{and} \quad B_i^* = \sum_{k=1}^{K} B_{ik}^{\otimes 2},$$

where $A_{ik} = 0$ for $i > P$ and $B_{ik} = 0$ for $i > Q$. A sufficient condition for the stationarity of the CBF model is given below, and works for other general distributions of Δ_t.

Theorem 1. *Suppose that* $\{\Delta_t\}$ *in model* (2.1) *is a sequence of i.i.d.* $n \times n$ *positive-definite random matrices with* $E\|\Delta_t\| < \infty$, *and*

(H1) the distribution of Δ_1, *denoted by* Γ, *is absolute continuous with respect to the Lebesgue measure;*

(H2) the point I_n *is in the interior of the support of* Γ*;*

(H3) $\rho\left(\sum_{i=1}^{M}(A_i^* + B_i^*)\right) < 1$.

Then, Y_t *in model* (2.1) *is strictly stationary, with* $E\|Y_t\| < \infty$. *Moreover,* Y_t *is positive Harris recurrent and geometrically ergodic.*

Remark 1. The results of Theorem 1 are similar to those in Boussama, Fuchs and Stelzer (2011), who study the stationarity of the BEKK model. Like Boussama, Fuchs and Stelzer (2011), the proof of Theorem 1 is based on the semi-polynomial Markov chains technique. However, it is relatively involved owing to the matrix nature of model (2.1).

As a special case, the results in Theorem 1 hold for the CAW model, in which Δ_t follows the Wishart distribution. Under conditions (H1) and (H2), condition (H3) is necessary and sufficient for the strict stationarity of Y_t, with $E\|Y_t\| < \infty$. However, the necessary and sufficient condition for the higher moments of Y_t is still unclear at this stage. Let K_{n^2} be the $n^2 \times n^2$ permutation matrix, such that $K_{n^2}vec(A) = vec(A')$ for any $n \times n$ matrix A. If $E\|Y_t\|^2 < \infty$, by similar arguments in Golosnoy, Gribisch and Liesenfeld (2012), we have the following:

(i) $\overline{y} := E\left(vec(Y_t)\right) = \left[I_{n^2} - \sum_{i=1}^{M}\left(A_i^* + B_i^*\right)\right]^{-1} vec(\Omega)$;

(ii) $vec\left[E\left(vec(Y_t)vec(Y_t)'\right)\right] = (\Pi + I_{n^4})\left(I_{n^4} - \sum_{i=1}^{\infty}\Phi_i^{\otimes 2}\Pi\right)^{-1} vec(\overline{y}) \otimes vec(\overline{y})$,

where $\Pi = \left[s_1(\nu) - 1\right] I_{n^4} + \left[s_2(\nu) I_{n^2} \otimes (I_{n^2} + K_{n^2})\right]\left[I_n \otimes K_{n^2} \otimes I_n\right]$, with

$$s_1(\nu) = \frac{(\nu_2 - n - 1)[\nu_1(\nu_2 - n - 2) + 2]}{\nu_1(\nu_2 - n)(\nu_2 - n - 3)}, \quad s_2(\nu) = \frac{(\nu_2 - n - 1)(\nu_1 + \nu_2 - n - 1)}{\nu_1(\nu_2 - n)(\nu_2 - n - 3)},$$

and $\Phi_0 = I_{n^2}$ and $\Phi_i = -B_i^* + \sum_{j=1}^{i}\left(A_j^* + B_j^*\right)\Phi_{i-j}$, for $i > 0$. Result (ii) clearly indicates that the parameters ν_1 and ν_2 affect the second moment of Y_t in a nonlinear way. Although a closed form of the third moment of Y_t is absent, similar effects from ν_1 and ν_2 are expected for the third moment of Y_t and, hence, the asymptotic distribution of the proposed estimator (see Theorem 3 below).

3. Maximum Likelihood Estimation

Let $\theta = (\gamma', \nu')' \in \Theta$ be the unknown parameter of model (2.1) with the true value $\theta_0 = (\gamma_0', \nu_0')'$, where $\Theta = \Theta_\gamma \times \Theta_\nu$ is the parametric space with $\Theta_\gamma \subset \mathbb{R}^{\tau_1}$ and $\Theta_\nu \subset \mathbb{R}^2$, $\gamma = (w', u')'$, $w = vech(\Omega)$, $u = (vec(A_{11})', \ldots, vec(A_{KP})', vec(B_{11})', \ldots, vec(B_{KQ})')$, and $\tau_1 = (1/2)n + [(P+Q)K + (1/2)]n^2$. Below, we assume that Θ_γ and Θ_ν are compact and θ_0 is an interior point of Θ.

Given the observations $\{Y_t\}_{t=1}^{T}$ and the initial values $\{Y_t\}_{t\le 0}$, the negative log-likelihood function based on (2.4) is

$$L(\theta) = \frac{1}{T}\sum_{t=1}^{T} l_t(\theta), \tag{3.1}$$

where

$$l_t(\theta) = \frac{\nu_1}{2}\log\left|\frac{\nu_2-n-1}{\nu_1}\Sigma_t(\gamma)\right| - \frac{\nu_1-n-1}{2}\log|Y_t| + \frac{\nu_1+\nu_2}{2}\log\left|I_n + \frac{\nu_1}{\nu_2-n-1}\Sigma_t^{-1}(\gamma)Y_t\right| + C(\nu),$$

with $C(\nu) = -\log\Lambda(\nu)$ and $\Sigma_t(\gamma)$ calculated recursively by

$$\Sigma_t(\gamma) = \Omega + \sum_{i=1}^{P}\sum_{k=1}^{K} A_{ki}Y_{t-i}A'_{ki} + \sum_{j=1}^{Q}\sum_{k=1}^{K} B_{kj}\Sigma_{t-j}(\gamma)B'_{kj}. \tag{3.2}$$

Clearly, $\Sigma_t(\gamma_0) = \Sigma_t$.

Because the initial values $\{Y_t\}_{t\le 0}$ are not observable, we modify $L(\theta)$ as

$$\widehat{L}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\widehat{l}_t(\theta), \tag{3.3}$$

where $\widehat{l}_t(\theta)$ is defined in the same way as $l_t(\theta)$, with $\Sigma_t(\gamma)$ replaced by $\widehat{\Sigma}_t(\gamma)$, and $\widehat{\Sigma}_t(\gamma)$ is calculated in the same way as $\Sigma_t(\gamma)$, based on a sequence of given constant matrices $h := \{Y_0, \ldots, Y_{-M+1}, \Sigma_0, \ldots, \Sigma_{-M+1}\}$. The minimizer, $\widehat{\theta} = (\widehat{\gamma}', \widehat{\nu}')'$, of $\widehat{L}(\theta)$ on Θ is called the MLE of θ_0. That is,

$$\widehat{\theta} = (\widehat{\gamma}', \widehat{\nu}')' = \operatorname*{argmin}_{\theta\in\Theta}\widehat{L}(\theta). \tag{3.4}$$

To study the asymptotic properties of $\widehat{\theta}$, we need two assumptions.

Assumption 1. *Y_t is strictly stationary and ergodic.*

Assumption 2. *For $\gamma \in \Theta_\gamma$, if $\gamma \neq \gamma_0$, $\Sigma_t(\gamma) \neq \Sigma_t(\gamma_0)$ almost surely (a.s.) for all t.*

Assumption 1 is standard. Assumption 2, which is in line with Comte and Lieberman (2003) and Hafner and Preminger (2009), is the identification condition. The following two theorems give the consistency and asymptotic normality of $\widehat{\theta}$, respectively.

Theorem 2. *Suppose that Assumptions 1–2 hold and $E\|Y_t\| < \infty$. Then, $\widehat{\theta} \xrightarrow{a.s.} \theta_0$ as $T \to \infty$.*

Theorem 3. *Suppose that Assumptions 1–2 hold, $E\|Y_t\|^3 < \infty$, and*

$$\mathcal{O} = E\left(\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\right) \text{ is invertible.} \tag{3.5}$$

Then, $\sqrt{T}(\widehat{\theta}-\theta_0) \xrightarrow{d} N(0, \mathcal{O}^{-1})$ *as* $T \to \infty$.

Based on the observations $\{Y_t\}_{t=1}^T$ and a sequence of given constant matrices h, we can use the analytic expression of $\partial^2 l_t(\theta)/\partial\theta\partial\theta'$ (see Appendix S4 in the Supplementary Material) to estimate $\mathcal{O}$ using its sample counterpart. As with the univariate ARCH-type models, the coefficients in the main diagonal line of Ω are positive to ensure the positive definiteness of Σ_t. Hence, the classical t or Wald test, which is constructed using the estimate of $\mathcal{O}$, cannot be used to detect whether their values are zeros; see Li et al. (2018) for more discussion on this context.

4. Model Diagnostic Checking

Diagnostic tests are crucial for model checking in multivariate time series analysis; see, for example, Li and McLeod (1981), Ling and Li (1997), Tse (2002), and many others. However, no tests exist for stationary matrix time series. In this section, we propose some new inner-product-based tests to check the adequacy of model (2.1).

Let $\mathfrak{z}_t(\gamma) = vec\big(\Sigma_t^{-1/2}(\gamma) Y_t \Sigma_t^{-1/2}(\gamma) - I_n\big)$ be the vectorized residual for a given γ, and let $\mathbf{b}_{t,j}(\gamma) = \mathfrak{z}_t'(\gamma)\mathfrak{z}_{t-j}(\gamma)$ be the inner product of two vectorized residuals at lag j. Then, we stack $\mathbf{b}_{t,j}(\gamma)$ up to lag l to construct $\mathcal{V}_l(\gamma)$, where

$$\mathcal{V}_l(\gamma) = \frac{1}{T}\sum_{t=l+1}^{T}\big(\mathbf{b}_{t,1}(\gamma), \mathbf{b}_{t,2}(\gamma), \ldots, \mathbf{b}_{t,l}(\gamma)\big)',$$

and $l \geq 1$ is a given integer. Our testing idea is motivated by the fact that if model (2.1) is adequate, $\mathfrak{z}_t(\gamma_0)$ is a sequence of i.i.d. random vectors with mean zero, and hence the value of $\mathcal{V}_l(\widehat{\gamma})$ is expected to be close to zero. To implement our test, we examine the asymptotic property of $\mathcal{V}_l(\widehat{\gamma})$ in the following theorem.

Theorem 4. *Suppose that Assumptions* 1–2 *hold,* $E\|Y_t\|^4 < \infty$, *and* (3.5) *holds. Then, if model* (2.1) *is correctly specified,* $\sqrt{T}\mathcal{V}_l(\widehat{\gamma}) \xrightarrow{d} N(0, \mathbf{V})$ *as* $T \to \infty$, *where* $\mathbf{V} = (I_l, \mathfrak{R}_1)\mathfrak{R}_2(I_l, \mathfrak{R}_1)'$ *with*

$$\mathfrak{R}_1 = E\begin{pmatrix} \mathfrak{z}_{t-1}'(\gamma_0)\left(\frac{\partial \mathfrak{z}_t(\gamma_0)}{\partial\theta'}\right) \\ \mathfrak{z}_{t-2}'(\gamma_0)\left(\frac{\partial \mathfrak{z}_t(\gamma_0)}{\partial\theta'}\right) \\ \vdots \\ \mathfrak{z}_{t-l}'(\gamma_0)\left(\frac{\partial \mathfrak{z}_t(\gamma_0)}{\partial\theta'}\right) \end{pmatrix} \times \mathcal{O}^{-1} \ \textit{and}\ \mathfrak{R}_2 = \begin{pmatrix} tr\{E^2[\mathfrak{z}_t'(\gamma_0)\mathfrak{z}_t(\gamma_0)]\}I_l & 0 \\ 0 & \mathcal{O} \end{pmatrix}.$$

Based on Theorem 4, we construct the inner-product-based test statistic

$$\Pi(l) = T\big[\mathcal{V}_l'(\widehat{\gamma})\widehat{\mathbf{V}}^{-1}\mathcal{V}_l(\widehat{\gamma})\big] \tag{4.1}$$

to detect the adequacy of model (2.1), where $\widehat{\mathbf{V}}$ is the sample counterpart of $\mathbf{V}$. If $\Pi(l)$ is larger than the upper-tailed critical value of $\chi^2(l)$, the fitted model (2.1) is not adequate at a given significance level. Otherwise, it is adequate.

Note that if we consider a test based on $\{\mathfrak{Z}_t(\widehat{\gamma})\}$ directly, the resulting limiting distribution is still chi-squared, but its degrees of freedom increase fast with the dimension n. To avoid this dilemma, we use the inner product of the residuals to construct our test $\Pi(l)$, the limiting distribution of which is independent of n. This new idea is different from the portmanteau test in Ling and Li (1997), in which the test statistic is constructed based on the auto-correlations of the transformed scale residuals. In our test, $\Pi(l)$ is based on the auto-covariances of the original vectorized residuals. Clearly, our idea can be extended easily to the framework in Ling and Li (1997). Our inner-product-based test $\Pi(l)$ takes the auto-covariances of all entries of $\mathfrak{Z}_t(\widehat{\gamma})$ into account, whereas the idea of a regression-based test in Tse (2002) considers only one entry of $\mathfrak{Z}_t(\widehat{\gamma})$ at a time. In view of this, we prefer to use the proposed inner-product idea for testing purposes.

5. Reduced CBF Models

Because the number of parameters in the CBF model is $O(n^2)$, the estimation of the CBF model may be computationally demanding when n is large. This section introduces two reduced CBF models that are feasible in fitting RCOV matrices with a large n.

5.1. The VT-CBF model

This subsection proposes a reduced CBF model by using the variance target (VT) technique in Engle and Mezrich (1996). This technique re-parameterizes the drift matrix Ω by using the theoretical mean of Y_t, such that the estimation of Ω is excluded in the implementation of the maximum likelihood estimation. Other related studies on VT time series models include those of Francq, Horváth and Zakoïan (2011) and Pedersen and Rahbek (2014).

To define our reduced model, we assume that Y_t is strictly stationary with a finite mean $S = E(Y_t)$. By taking the expectation on both sides of (2.3), we have

$$\Omega = S - \sum_{i=1}^{P}\sum_{k=1}^{K} A_{ki}SA_{ki}' - \sum_{j=1}^{Q}\sum_{k=1}^{K} B_{kj}SB_{kj}', \tag{5.1}$$

because $S = E(Y_t) = E(\Sigma_t)$. With the help of (5.1), model (2.1) becomes

$$Y_t = \Sigma_t^{1/2} \Delta_t \Sigma_t^{1/2}, \tag{5.2}$$

where all notation is inherited from model (2.1), except that

$$\begin{aligned} \Sigma_t = S &- \sum_{i=1}^{P} \sum_{k=1}^{K} A_{ki} S A'_{ki} - \sum_{j=1}^{Q} \sum_{k=1}^{K} B_{kj} S B'_{kj} \\ &+ \sum_{i=1}^{P} \sum_{k=1}^{K} A_{ki} Y_{t-i} A'_{ki} + \sum_{j=1}^{Q} \sum_{k=1}^{K} B_{kj} \Sigma_{t-j} B'_{kj}. \end{aligned} \tag{5.3}$$

We call model (5.2) the VT-CBF model. Clearly, this reduced model shares the same probabilistic properties as the full CBF model. Although the VT-CBF model has the same number of parameters as the full CBF model, its two-step estimator, given below, is computationally easier than the MLE for the full CBF model.

To present this two-step estimator, we let $\theta_v = (\delta', \nu')' \in \Theta_v$ be the unknown parameters of model (5.2), and let its true value be $\theta_{v0} = (\delta_0', \nu_0')'$, where $\Theta_v = \Theta_\delta \times \Theta_\nu$ is the parametric space with $\Theta_\delta = \Theta_s \times \Theta_u \subset \mathbb{R}^{\tau_2}$, $\tau_2 = [(P+Q)K+1]n^2$, and $\Theta_\nu \subset \mathbb{R}^2$. Let $\delta = (s', u')'$ with $s = vec(S)$, $\Theta_s \in \mathbb{R}^{n^2}$, and $\Theta_u \in \mathbb{R}^{[(P+Q)Kn^2]}$. As before, we assume that Θ_δ and Θ_ν are compact, and that θ_{v0} is an interior point of Θ_v.

In the first step, we estimate s by $\widehat{s}_v$, where $\widehat{s}_v = vec\left(\overline{Y_t}\right) := vec((1/T)\sum_{t=1}^{T} Y_t)$. In the second step, we estimate the remaining parameters $\zeta = (u', \nu')'$ by the constrained MLE based on the following modified log-likelihood function:

$$\widehat{L}_v(\theta_v) = \frac{1}{T} \sum_{t=1}^{T} \widehat{l}_{vt}(\theta_v), \tag{5.4}$$

where

$$\begin{aligned} \widehat{l}_{vt}(\theta_v) = \frac{\nu_1}{2} \log \left| \frac{\nu_2 - n - 1}{\nu_1} \widehat{\Sigma}_{vt}(\delta) \right| &- \frac{\nu_1 - n - 1}{2} \log |Y_t| \\ &+ \frac{\nu_1 + \nu_2}{2} \log \left| I_n + \frac{\nu_1}{\nu_2 - n - 1} \widehat{\Sigma}_{vt}^{-1}(\delta) Y_t \right| + C(\nu), \end{aligned}$$

and $\widehat{\Sigma}_{vt}(\delta)$ is calculated recursively by

$$\widehat{\Sigma}_{vt}(\delta) = S - \sum_{i=1}^{P} \sum_{k=1}^{K} A_{ki} S A'_{ki} - \sum_{j=1}^{Q} \sum_{k=1}^{K} B_{kj} S B'_{kj}$$

$$+\sum_{i=1}^{P}\sum_{k=1}^{K}A_{ki}Y_{t-i}A'_{ki}+\sum_{j=1}^{Q}\sum_{k=1}^{K}B_{kj}\widehat{\Sigma}_{vt-j}(\delta)B'_{kj}, \tag{5.5}$$

based on a sequence of given constant matrices h. Clearly, $\widehat{L}_v(\theta_v)$ is analogous to $\widehat{L}(\theta)$ in (3.3), and is a modification of the following log-likelihood function:

$$L_v(\theta_v)=\frac{1}{T}\sum_{t=1}^{T}l_{vt}(\theta_v), \tag{5.6}$$

where $l_{vt}(\theta_v)$ is defined in the same way as $\widehat{l}_{vt}(\theta_v)$, with $\widehat{\Sigma}_{vt}(\delta)$ replaced by $\Sigma_{vt}(\delta)$, and $\Sigma_{vt}(\delta)$ is calculated recursively by

$$\begin{aligned}\Sigma_{vt}(\delta)=S&-\sum_{i=1}^{P}\sum_{k=1}^{K}A_{ki}SA'_{ki}-\sum_{j=1}^{Q}\sum_{k=1}^{K}B_{kj}SB'_{kj}\\&+\sum_{i=1}^{P}\sum_{k=1}^{K}A_{ki}Y_{t-i}A'_{ki}+\sum_{j=1}^{Q}\sum_{k=1}^{K}B_{kj}\Sigma_{vt-j}(\delta)B'_{kj},\end{aligned} \tag{5.7}$$

based on the observations $\{Y_t\}_{t=1}^{T}$ and the initial values $\{Y_t\}_{t\le 0}$. The minimizer, $\widehat{\zeta}_v=(\widehat{u}'_v,\widehat{\nu}'_v)'$, of $\widehat{L}_v(\widehat{s}_v,\zeta)$ on $\Theta_u\times\Theta_\nu$ is the constrained MLE of $(u'_0,\nu'_0)'$. That is,

$$(\widehat{u}'_v,\widehat{\nu}'_v)'=\operatorname*{argmin}_{\zeta\in\Theta_u\times\Theta_\nu}\ \widehat{L}_v(\widehat{s}_v,\zeta). \tag{5.8}$$

Now, we call $\widehat{\theta}_v=(\widehat{s}'_v,\widehat{\zeta}'_v)'$ the two-step estimator of θ_v in model (5.2). Let $\Psi(u)=\big(I_{n^2}-\sum_{i=1}^{M}A_i^*-\sum_{i=1}^{M}B_i^*\big)^{-1}\big(I_{n^2}-\sum_{i=1}^{M}B_i^*\big)$ and $w_t(\theta_v)=\begin{pmatrix}\Psi(u)vec(Y_t-\Sigma_{vt}(\delta))\\ \partial l_{vt}(\theta_v)/\partial\zeta\end{pmatrix}$. The following two theorems give the consistency and asymptotic normality of $\widehat{\theta}_v$, respectively.

Theorem 5. *Suppose that Assumptions* 1–2 *hold and* $E\|Y_t\|<\infty$. *Then,* $\widehat{\theta}_v\xrightarrow{a.s.}\theta_{v0}$ *as* $T\to\infty$.

Theorem 6. *Suppose that Assumptions* 1–2 *hold,* $E\|Y_t\|^3<\infty$, *and*

$$J_1=E\left[\frac{\partial^2 l_{vt}(\theta_{v0})}{\partial\zeta\partial\zeta'}\right]\ \textit{is invertible.} \tag{5.9}$$

Then, $\sqrt{T}(\widehat{\theta}_v-\theta_{v0})\xrightarrow{d}N(0,\mathcal{O}_v)$ *as* $T\to\infty$, *where*

$$\mathcal{O}_v = \begin{pmatrix} I_{n^2} & 0 \\ -J_1^{-1}J_2 & -J_1^{-1} \end{pmatrix} E(w_t w_t') \begin{pmatrix} I_{n^2} & 0 \\ -J_1^{-1}J_2 & -J_1^{-1} \end{pmatrix}',$$

with $J_2 = E[\partial^2 l_{vt}(\theta_{v0})/\partial\zeta\partial s']$ *and* $w_t = w_t(\theta_{v0})$.

As before, we can use the sample counterparts of the analytic expressions of $\partial l_{vt}(\theta_v)/\partial\theta_v$ and $\partial^2 l_{vt}(\theta_v)/\partial\theta_v\partial\theta_v'$ to estimate $\mathcal{O}_v$. Although the VT-CBF model can be estimated using the aforementioned two-step estimation procedure, it still has to handle a large number of estimated parameters, with order $O(n^2)$, caused by the parameter matrices A_{ki} and B_{kj}. To construct a more parsimonious VT-CBF model, we impose some restrictions on A_{ki} and B_{kj}. McCurdy and Stengos (1992) and Engle and Kroner (1995) have suggested using diagonal volatility models, which not only avoid over-parameterization, but also reflect the fact that the variances and the covariances rely more on their own past than they do on the history of other variances or covariances. Motivated by this, we assume that all A_{ki} and B_{kj} have a diagonal structure, leading to a diagonal VT-CBF model. Clearly, the number of estimated parameters in the diagonal VT-CBF model is $O(n)$, which is feasible for a moderately large, but fixed n.

Next, similarly to $\Pi(l)$ in (4.1), we construct inner-product-based test statistics to check the adequacy of model (2.1) based on the two-step estimator $\widehat{\theta}_v$. Let $\delta_0 = (s_0', u_0')'$, $\widehat{\delta}_v = (\widehat{s}_v', \widehat{u}_v')'$, $\mathfrak{z}_{vt}(\delta) = vec(\Sigma_{vt}^{-1/2}(\delta)Y_t\Sigma_{vt}^{-1/2}(\delta) - I_n)$ be the residual vector for a given δ, $\mathbf{b}_{vt,j}(\delta) = \mathfrak{z}_{vt}'(\delta)\mathfrak{z}_{vt-j}(\delta)$ be the inner product of the residuals at lag j, and

$$\mathcal{V}_{vl}(\delta) = \frac{1}{T}\sum_{t=l+1}^{T} \big(\mathbf{b}_{vt,1}(\delta), \mathbf{b}_{vt,2}(\delta), \ldots, \mathbf{b}_{vt,l}(\delta)\big)'.$$

The asymptotic property of $\mathcal{V}_{vl}(\widehat{\delta}_v)$ is given in the following theorem.

Theorem 7. *Suppose that Assumptions* 1–2 *hold,* $E\|Y_t\|^4 < \infty$*, and* (5.9) *holds. Then, if model* (2.1) *is correctly specified,* $\sqrt{T}\mathcal{V}_{vl}(\widehat{\delta}_v) \xrightarrow{d} N(0, \mathbf{V}_v)$ *as* $T \to \infty$*, where* $\mathbf{V}_v = (I_l, \mathfrak{R}_{1v})\mathfrak{R}_{2v}(I_l, \mathfrak{R}_{1v})'$*, with*

$$\mathfrak{R}_{1v} = E\begin{pmatrix} \mathfrak{z}_{vt-1}'(\delta_0)\left(\frac{\partial\mathfrak{z}_{vt}(\delta_0)}{\partial\theta'}\right) \\ \mathfrak{z}_{vt-2}'(\delta_0)\left(\frac{\partial\mathfrak{z}_{vt}(\delta_0)}{\partial\theta'}\right) \\ \vdots \\ \mathfrak{z}_{vt-l}'(\delta_0)\left(\frac{\partial\mathfrak{z}_{vt}(\delta_0)}{\partial\theta'}\right) \end{pmatrix} \times \begin{pmatrix} I_{n^2} & 0 \\ -J_1^{-1}J_2 & -J_1^{-1} \end{pmatrix}$$

and

$$\mathfrak{R}_{2v} = \begin{pmatrix} tr\{E^2[\mathfrak{Z}_{vt}(\delta_0)'\mathfrak{Z}_{vt}(\delta_0)]\}I_l & 0 \\ 0 & E(w_t w_t') \end{pmatrix}.$$

By the preceding theorem, we can adopt the test statistic

$$\Pi_v(l) = T[\mathcal{V}_{vl}'(\widehat{\delta}_v)\widehat{\mathbf{V}}_v^{-1}\mathcal{V}_{vl}(\widehat{\delta}_v)] \tag{5.10}$$

to detect the adequacy of model (2.1), where $\widehat{\mathbf{V}}_v$ is the sample counterpart of $\mathbf{V}_v$. If $\Pi_v(l)$ is larger than the upper-tailed critical value of $\chi^2(l)$ at a given significance level, the fitted model (2.1) is not adequate. Otherwise, it is adequate.

5.2. The factor CBF model

In modern data analysis, the dimension n may grow with the sample size T in many cases, making the CBF (or VT-CBF) models computationally infeasible. In addition, the dimension n may be proportional to m (the average intra-day sample size across all assets and all days), in which case, the methods used to calculate Y_t for fixed n deliver an inconsistent estimator of Y_t^*; see, for example, Wang and Zou (2010) and Tao et al. (2011) for surveys. To overcome this difficulty, we use the thresholding average realized volatility matrix (TARVM) estimator of Tao et al. (2011) to calculate Y_t. The TARVM is based on the ARVM (Wang and Zou (2010)), which is estimated by taking the average of the constructed realized volatility matrices according to different predetermined sampling frequencies. The TARVM further thresholds the elements in each estimated RCOV matrix from the ARVM method, so that a certain sparsity structure is retained and the resulting estimator is consistent for large n, which can be growing with (or even larger than) T. For more recent works in this direction, refer to Aït-Sahalia and Xiu (2017) and Kim et al. (2018), and the references therein.

Because the dimension of Y_t may be very large, it seems difficult to study the dynamics of Y_t without imposing a specific structure. Here, we adopt the factor model proposed by Tao et al. (2011) by assuming that

$$Y_t^* = FY_{ft}^*F' + Y_0^*, \tag{5.11}$$

where Y_{ft}^* is an $r \times r$ positive-definite factor covariance matrix, with r being a fixed integer (much smaller than n), Y_0^* is an $n \times n$ positive-definite constant matrix, and F is an $n \times r$ factor loading matrix normalized by the constraint $F'F = I_r$. In model (5.11), the dynamic structure of Y_t^* is driven by that of a lower-dimensional latent process Y_{ft}^*, while Y_0^* represents the static part of Y_t^*.

In (5.11), only the column space of F can be identified, and F is not identified even if $F'F = I_r$ is imposed. This is because Y_t^* is unchanged when F and Y_{ft}^* are replaced by $F_\dagger = FR$ and $Y_{ft,\dagger}^* = R^{-1}Y_{ft}^*R^{-1\prime}$, respectively, when R is any $r \times r$ matrix satisfying $R'R = I_r$.

Define

$$\overline{Y}^* = \frac{1}{T}\sum_{t=1}^{T} Y_t^*, \quad \overline{S}^* = \frac{1}{T}\sum_{t=1}^{T} \left\{Y_t^* - \overline{Y}^*\right\}^2,$$

and

$$\overline{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t, \quad \overline{S} = \frac{1}{T}\sum_{t=1}^{T} \left\{Y_t - \overline{Y}\right\}^2.$$

Then, we estimate Y_{ft}^*, Y_0^*, and F by

$$\widehat{Y}_{ft} = \widehat{F}'Y_t\widehat{F}, \quad \widehat{Y}_0^* = \overline{Y} - \widehat{F}\widehat{F}'\overline{Y}\widehat{F}\widehat{F}', \quad \text{and} \quad \widehat{F} = (\widehat{f}_1, \ldots, \widehat{f}_r), \tag{5.12}$$

respectively, where $\widehat{f}_1, \ldots, \widehat{f}_r$ are the eigenvectors of $\overline{S}$ corresponding to its r largest eigenvalues. As suggested by Lam and Yao (2012) and Ahn and Horenstein (2013), we may select r such that the r largest ratios of adjacent eigenvalues are significantly larger.

In order to study the asymptotics of the proposed estimators, we introduce the following technical assumptions.

Assumption 3. *All row vectors of F' and Y_0^* satisfy the sparsity condition below. For an n-dimensional vector $(x_1, \ldots, x_n)$, we say it is sparse if it satisfies*

$$\sum_{i=1}^{n} |x_i|^{\delta_*} \le U\pi(n),$$

where $\delta_* \in [0, 1)$, U is a positive constant, and $\pi(n)$ is a deterministic function of n that grows slowly in n, with typical examples $\pi(n) = 1$ or $\log(n)$.

Assumption 4. *The factor model (5.11) has r fixed factors, and the matrices Y_0^* and Y_{ft}^* satisfy $\|Y_0^*\| < \infty$ and $\max_{1\le t\le T} \|Y_{ft,jj}^*\| = O_p(B(T))$ for $j = 1, 2, \ldots, r$, where $Y_{ft,jj}^*$ is the jth diagonal entry of Y_{ft}^*, and $1 \le B(T) = o(T)$.*

Assumption 5. $\max_{1\le t\le T} \|Y_t^* - Y_t\| = O_p(A(n, m, T))$ *for some rate function $A(n, m, T)$, such that $A(n, m, T)B^5(T) = o(1)$.*

Assumptions 3–5 are sufficient to prove the consistency of $\widehat{Y}_{ft}$. For the TARVM, we can take $A(n, m, T) = \pi(n)[e_m(n^2T)^{1/\beta}]^{1-\delta_*} \log T$ and $B(T) = \log T$ with $e_m = m^{-1/6}$, such that $A(n, m, T)B^5(T) = o(1)$ for large β; see Tao et al. (2011). Note that Assumptions 3–5 do not rule out the case that n is larger than

T, as long as n^2T grows more slowly than $m^{\beta/6}$. For other estimators, the rate $A(n,m,T)$ may be improved; see Tao, Wang and Zhou (2013).

Theorem 8. *Suppose that Assumptions* 3–5 *and the conditions in Theorem* 3 *hold. Then, as* n, m, *and* T *go to infinity,*

(i) $F'\widehat{F} - I_r = O_p(A(n,m,T)B(T))$,

(ii) $\widehat{Y}_{ft} - Y_{ft} = O_p(A^{1/2}(n,m,T)B^{3/2}(T))$,

where $Y_{ft} = Y_{ft}^* + F'Y_0^*F$ *and* $F = (f_1, \ldots, f_r)$, *with* $f_1, \ldots, f_r$ *being the eigenvectors of* $\bar{S}^*$ *corresponding to its* r *largest eigenvalues.*

The above theorem indicates that $\widehat{Y}_{ft}$ is a more consistent estimator of Y_{ft} than is Y_{ft}^*. Next, we assume that Y_{ft} satisfies the CBF model; that is,

$$Y_{ft}|\mathcal{G}_{t-1} \sim F\left(\nu, \frac{\nu_2 - n - 1}{\nu_1}\Sigma_{ft}\right), \tag{5.13}$$

with $E(Y_{ft}|\mathcal{G}_{t-1}) = \Sigma_{ft}$, where Σ_{ft} is defined in the same way as Σ_t in (2.3), with Y_t replaced by Y_{ft}, and the remaining notation and setup inherited from model (2.1). We call models (5.11) and (5.13) the factor CBF (F-CBF) model. In particular, if Σ_{ft} has the HAR dynamical structure in (2.6), the resulting model is called the factor CBF-HAR (F-CBF-HAR) model. Based on this model, we have $Y_t^* = F(Y_{ft} - F'Y_0^*F)F' + Y_0^*$. Because $Y_t \approx Y_t^*$, this implies that we can study the large-dimensional matrix Y_t by using an $r \times r$ low-dimensional matrix Y_{ft}.

Because Y_{ft} is not observable, we should estimate model (5.13) based on $\widehat{Y}_{ft}$. Hence, we consider a feasible MLE of θ_0 in model (5.13) given by

$$\widehat{\theta}_{1f} = (\widehat{\gamma}_{1f}', \widehat{\nu}_{1f}')' = \underset{\theta \in \Theta}{\operatorname{argmin}}\ \widehat{L}_f(\theta),$$

where $\widehat{L}_f(\theta)$ is defined in the same way as $\widehat{L}(\theta)$ in (3.3), with Y_t and $\widehat{\Sigma}_t(\gamma)$ replaced by $\widehat{Y}_{ft}$ and $\widehat{\Sigma}_{ft}(\gamma)$, respectively. The following theorem shows that $\widehat{\theta}_{1f}$ is consistent with the ideal MLE $\widehat{\theta}_{2f}$ based on Y_{ft}, where

$$\widehat{\theta}_{2f} = (\widehat{\gamma}_{2f}', \widehat{\nu}_{2f}')' = \underset{\theta \in \Theta}{\operatorname{argmin}}\ L_f(\theta),$$

and $L_f(\theta)$ is defined in the same way as $L(\theta)$ in (3.1), with Y_t and $\Sigma_t(\gamma)$ replaced by Y_{ft} and $\Sigma_{ft}(\gamma)$, respectively.

Theorem 9. *Suppose that the conditions in Theorem* 8 *hold. Then, as* n, m, *and* T *go to infinity,* $\widehat{\theta}_{1f} - \widehat{\theta}_{2f} = O_p(B(T)/T) + O_p(A^{1/2}(n,m,T)B^{5/2}(T))$.

Because the dimension of Y_{ft} is r (much smaller than n), the calculation of $\widehat{\theta}_{1f}$ is computationally feasible. In order to further reduce the number of parameters in model (5.13), we can also assume that Y_{ft} follows a VT-CBF model. This leads to the F-VT-CBF model, which includes the F-VT-CBF-HAR model as a special case. For this F-VT-CBF model, we consider its feasible two-step estimator $\widehat{\theta}_{1fv} = (\widehat{s}'_{1fv}, \widehat{\zeta}'_{1fv})'$, where

$$\widehat{s}_{1fv} = \frac{1}{T}\sum_{t=1}^{T}\widehat{Y}_{ft}, \quad \widehat{\zeta}_{1fv} = (\widehat{u}'_{1fv}, \widehat{\nu}'_{1fv})' = \underset{\zeta\in\Theta_u\times\Theta_\nu}{\operatorname{argmin}}\ \widehat{L}_{fv}(\widehat{s}_{1fv}, \zeta),$$

and $\widehat{L}_{fv}(\theta_v)$ is defined in the same way as $\widehat{L}_v(\theta_v)$ in (5.4), with Y_t and $\widehat{\Sigma}_{vt}(\delta)$ replaced by $\widehat{Y}_{ft}$ and $\widehat{\Sigma}_{fvt}(\delta)$, respectively. Similarly to Theorem 9, $\widehat{\theta}_{1fv}$ is consistent with the ideal two-step estimator $\widehat{\theta}_{2fv} = (\widehat{s}'_{2fv}, \widehat{\zeta}'_{2fv})'$ based on Y_{ft}, where

$$\widehat{s}_{2fv} = \frac{1}{T}\sum_{t=1}^{T}Y_{ft}, \quad \widehat{\zeta}_{2fv} = (\widehat{u}'_{2fv}, \widehat{\nu}'_{2fv})' = \underset{\zeta\in\Theta_u\times\Theta_\nu}{\operatorname{argmin}}\ L_{fv}(\widehat{s}_{2fv}, \zeta),$$

and $L_{fv}(\theta_v)$ is defined in the same way as $L(\theta_v)$ in (5.6), with Y_t and $\Sigma_t(\delta)$ replaced by Y_{ft} and $\Sigma_{fvt}(\delta)$, respectively.

Theorem 10. *Suppose that the conditions in Theorem* 8 *hold. Then, as n, m, and T go to infinity,*

(i) $\widehat{s}_{1fv} - \widehat{s}_{2fv} = O_p(A^{1/2}(n, m, T)B^{3/2}(T))$,

(ii) $\widehat{\zeta}_{1fv} - \widehat{\zeta}_{2fv} = O_p(B(T)/T) + O_p(A^{1/2}(n, m, T)B^{5/2}(T))$.

In particular, if Y_{ft} follows a diagonal VT-CBF model, the number of estimated parameters in model (5.13) is $O(r)$, which is easy to calculate in practice. In view of model (5.11) and the fact that $F'F = I_r$, we can predict Y_t by either $\widehat{F}\widehat{\Sigma}_{ft}(\widehat{\gamma}_{1f})\widehat{F}' + \widehat{Y}_0^*$ based on $\widehat{\theta}_{1f}$ or by $\widehat{F}\widehat{\Sigma}_{fvt}(\widehat{\delta}_{1fv})\widehat{F}' + \widehat{Y}_0^*$ based on $\widehat{\theta}_{1fv}$, where $\widehat{\delta}_{1fv} = (\widehat{s}'_{1fv}, \widehat{u}'_{1fv})'$.

6. Simulation

In this section, we first assess the performance of the MLE $\widehat{\theta}$ and the two-step estimator $\widehat{\theta}_v$ in the finite sample. We generate 1,000 replications of sample size $T = 1{,}000$ and 2,000 from the following model:

$$Y_t = \Sigma_t^{1/2}\Delta_t\Sigma_t^{1/2} \text{ with } \Sigma_t = \Omega_0 + A_{10}Y_{t-1}A'_{10} + B_{10}\Sigma_{t-1}B'_{10}, \tag{6.1}$$

where

$$\Omega_0 = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.2 & 0.5 & 0.25 \\ 0.3 & 0.25 & 0.5 \end{pmatrix}, \ A_{10} = \begin{pmatrix} 0.4 & 0 & 0 \\ 0 & 0.55 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}, \ B_{10} = \begin{pmatrix} 0.4 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0.5 \end{pmatrix},$$

$\{\Delta_t\}$ is a sequence of independent $F\big(\nu_0, ((\nu_{20} - n - 1)/\nu_{10})I_n\big)$ distributed random matrices with $n = 3$, and $\nu_0 = (10, 8), (15, 10)$, or $(20, 10)$. For each repetition, we calculate $\widehat{\theta}$, $\widehat{\theta}_v$, and their related asymptotic standard deviations. For $\widehat{\theta}_v$, we report the results related to Ω instead of S, and hence the asymptotic standard deviation of the estimated parameters in Ω is absent in this case.

Table 1 reports the sample bias, sample standard deviation (SD), and average asymptotic standard deviation (AD) of $\widehat{\theta}$ and $\widehat{\theta}_v$. From this table, we can see that the biases of both estimators are small relative to the magnitude of the parameters, and they become smaller as the sample size T increases. This ensures the accuracy of both estimators. Furthermore, we find that the SDs are, in general, close to the ADs for both estimators, and all of the SDs and ADs become smaller as T increases from 1,000 to 2,000. In terms of ADs or SDs, $\widehat{\theta}$ is, in general, more efficient than $\widehat{\theta}_v$, although this efficiency advantage is weak for many parameters. However, the estimation time for $\widehat{\theta}_v$ is almost 70% of that for $\widehat{\theta}$, and this computation advantage can be more significant when n increases.

Next, we examine the performance of the inner-product-based tests $\Pi(l)$ and $\Pi_v(l)$ in the finite sample. We generate 1,000 replications of sample size $T = 1{,}000$ and 2,000 from the following model:

$$Y_t = \Sigma_t^{1/2}\Delta_t\Sigma_t^{1/2} \text{ with } \Sigma_t = \Omega_0 + A_{10}Y_{t-1}A_{10}' + A_{20}Y_{t-2}A_{20}' + B_{10}\Sigma_{t-1}B_{10}', \tag{6.2}$$

where the values of Ω_0, A_{10}, and B_{10} are chosen as in (6.1), $A_{20} = \text{diag}\{\lambda, \lambda, \lambda\}$ is a diagonal matrix with $\lambda = 0, 0.05, 0.1, 0.15, 0.2$, and $\{\Delta_t\}$ is a sequence of independent $F\big(\nu_0, ((\nu_{20} - n - 1)/\nu_{10})I_n\big)$ distributed random matrices with $n = 3$ and $\nu_0 = (10, 8)$. We fit each replication using the CBF model with $(K, P, Q) = (1, 1, 1)$, and use $\Pi(l)$ and $\Pi_v(l)$ to check whether the fitted model is adequate. Here, we set the significance level $\alpha = 0.05$ and $l = 2, 3, 4, 5, 6$. The empirical sizes and power of both tests are reported in Table 2, with sizes corresponding to the results for the case of $\lambda = 0$. From Table 2, we find that $\Pi(l)$ and $\Pi_v(l)$ always have accurate sizes, although they are slightly oversized for small T. The power of both test is as expected. First, all of the power values become larger as T increases. Second, both tests become more powerful as λ becomes larger.

Table 1. The results of the MLE $\widehat{\theta}$ and two-step estimator $\widehat{\theta}_v$ for model (6.1).

	T			ν_1	ν_2	$A_{1,11}$	$A_{1,22}$	$A_{1,33}$	$B_{1,11}$	$B_{1,22}$	$B_{1,33}$	Ω_{11}	Ω_{21}	Ω_{31}	Ω_{22}	Ω_{32}	Ω_{33}
Case 1	1,000	$\widehat{\theta}$	Bias	0.0320	0.0160	-0.0014	-0.0029	-0.0009	-0.0151	-0.0112	-0.0102	-0.0005	0.0028	0.0057	-0.0009	0.0037	0.0053
			ESD	0.3914	0.2452	0.0255	0.0249	0.0240	0.1170	0.0964	0.0728	0.0600	0.0188	0.0337	0.0419	0.0248	0.0601
			ASD	0.4111	0.2563	0.0258	0.0259	0.0241	0.1103	0.0892	0.0652	0.0586	0.0179	0.0323	0.0402	0.0232	0.0562
		$\widehat{\theta}_v$	Bias	-0.0080	0.0382	-0.0005	-0.0030	0.0000	-0.0130	-0.0088	-0.0096	-0.0020	0.0020	0.0047	-0.0030	0.0033	0.0040
			ESD	0.3884	0.2607	0.0263	0.0272	0.0255	0.1165	0.0956	0.0728	0.0614	0.0229	0.0366	0.0433	0.0291	0.0615
			ASD	0.4024	0.2619	0.0266	0.0282	0.0258	0.1207	0.1046	0.0742						
	2,000	$\widehat{\theta}$	Bias	0.0188	0.0072	-0.0003	-0.0018	-0.0002	-0.0111	-0.0036	-0.0034	0.0019	0.0014	0.0030	-0.0008	0.0012	0.0014
			ESD	0.2767	0.1733	0.0174	0.0179	0.0168	0.0797	0.0633	0.0459	0.0431	0.0130	0.0231	0.0293	0.0169	0.0405
			ASD	0.2880	0.1797	0.0181	0.0182	0.0169	0.0767	0.0615	0.0447	0.0417	0.0124	0.0226	0.0287	0.0162	0.0395
		$\widehat{\theta}_v$	Bias	-0.0020	0.0196	0.0002	-0.0020	0.0003	-0.0103	-0.0024	-0.0030	0.0010	0.0007	0.0023	-0.0020	0.0008	0.0005
			ESD	0.2767	0.1876	0.0181	0.0194	0.0179	0.0800	0.0633	0.0458	0.0438	0.0161	0.0250	0.0304	0.0200	0.0416
			ASD	0.2856	0.1871	0.0187	0.0198	0.0180	0.0823	0.0640	0.0458						
Case 2	1,000	$\widehat{\theta}$	Bias	0.0900	0.0132	-0.0017	-0.0021	-0.0010	-0.0159	-0.0075	-0.0092	-0.0023	0.0026	0.0049	-0.0022	0.0031	0.0053
			ESD	0.8099	0.3597	0.0240	0.0245	0.0227	0.1242	0.0974	0.0690	0.0626	0.0177	0.0340	0.0418	0.0237	0.0591
			ASD	0.8413	0.3598	0.0250	0.0243	0.0227	0.1158	0.0921	0.0672	0.0602	0.0175	0.0328	0.0410	0.0230	0.0578
		$\widehat{\theta}_v$	Bias	0.0255	0.0353	-0.0011	-0.0019	-0.0009	-0.0154	-0.0067	-0.0089	-0.0027	0.0025	0.0041	-0.0031	0.0023	0.0041
			ESD	0.7985	0.3659	0.0244	0.0259	0.0232	0.1239	0.0974	0.0689	0.0632	0.0200	0.0352	0.0430	0.0257	0.0593
			ASD	0.8162	0.3585	0.0252	0.0253	0.0234	0.1290	0.1034	0.0724						
	2,000	$\widehat{\theta}$	Bias	0.0702	0.0000	-0.0005	-0.0011	0.0000	-0.0074	-0.0021	-0.0041	-0.0008	0.0009	0.0020	-0.0017	0.0010	0.0020
			ESD	0.5912	0.2515	0.0173	0.0174	0.0158	0.0801	0.0665	0.0467	0.0434	0.0123	0.0228	0.0299	0.0163	0.0412
			ASD	0.5871	0.2517	0.0175	0.0171	0.0159	0.0805	0.0640	0.0461	0.0437	0.0121	0.0230	0.0293	0.0161	0.0406
		$\widehat{\theta}_v$	Bias	0.0384	0.0112	-0.0002	-0.0009	-0.0001	-0.0071	-0.0018	-0.0040	-0.0010	-0.0007	0.0016	-0.0021	0.0005	0.0012
			ESD	0.5874	0.2532	0.0175	0.0182	0.0163	0.0800	0.0664	0.0467	0.0437	0.0138	0.0235	0.0305	0.0176	0.0413
			ASD	0.5792	0.2537	0.0177	0.0178	0.0163	0.0827	0.0668	0.0472						
Case 3	1,000	$\widehat{\theta}$	Bias	0.1521	0.0294	-0.0013	-0.0021	-0.0007	-0.0165	-0.0092	-0.0082	-0.0023	0.0026	0.0046	-0.0026	0.0032	0.0031
			ESD	1.4019	0.3340	0.0237	0.0237	0.0213	0.1253	0.0979	0.0712	0.0615	0.0173	0.0353	0.0418	0.0231	0.0599
			ASD	1.4496	0.3442	0.0242	0.0235	0.0220	0.1127	0.0904	0.0654	0.0586	0.0169	0.0319	0.0399	0.0224	0.0463
		$\widehat{\theta}_v$	Bias	0.0433	0.0475	-0.0008	-0.0013	-0.0006	-0.0156	-0.0083	-0.0082	-0.0031	0.0025	0.0040	-0.0031	0.0026	0.0021
			ESD	1.3774	0.3416	0.0239	0.0246	0.0220	0.1250	0.0975	0.0725	0.0614	0.0190	0.0358	0.0423	0.0250	0.0605
			ASD	1.4084	0.3443	0.0248	0.0246	0.0227	0.1409	0.0996	0.0755						
	2,000	$\widehat{\theta}$	Bias	0.0737	0.0190	-0.0006	-0.0009	0.0001	-0.0061	-0.0047	-0.0052	-0.0016	0.0010	0.0018	-0.0010	0.0016	0.0022
			ESD	1.0087	0.2469	0.0169	0.0163	0.0149	0.0794	0.0671	0.0480	0.0418	0.0118	0.0227	0.0295	0.0161	0.0418
			ASD	1.0057	0.2411	0.0170	0.0165	0.0155	0.0787	0.0630	0.0453	0.0429	0.0117	0.0225	0.0286	0.0157	0.0397
		$\widehat{\theta}_v$	Bias	0.0192	0.0286	-0.0004	-0.0004	0.0000	-0.0058	-0.0045	-0.0051	-0.0020	0.0008	0.0013	-0.0012	0.0011	0.0015
			ESD	1.0022	0.2511	0.0170	0.0171	0.0153	0.0795	0.0673	0.0480	0.0419	0.0131	0.0231	0.0300	0.0176	0.0418
			ASD	0.9923	0.2434	0.0172	0.0173	0.0158	0.0817	0.0652	0.0462						

† Cases 1–3 correspond to $\nu_0 = (10, 8), (15, 10)$, and $(20, 10)$, respectively.

Table 2. The results of $\Pi(l)$ and $\Pi_v(l)$ for model (6.2).

		$l=2$		$l=3$		$l=4$		$l=5$		$l=6$	
λ	T	$\Pi(l)$	$\Pi_v(l)$	$\Pi(l)$	$\Pi_v(l)$	$\Pi(l)$	$\Pi_v(l)$	$\Pi(l)$	$\Pi_v(l)$	$\Pi(l)$	$\Pi_v(l)$
0	1,000	0.043	0.037	0.048	0.045	0.052	0.054	0.047	0.048	0.049	0.054
	2,000	0.048	0.056	0.058	0.059	0.053	0.054	0.052	0.059	0.051	0.052
0.05	1,000	0.048	0.045	0.051	0.048	0.058	0.053	0.060	0.052	0.061	0.062
	2,000	0.060	0.063	0.063	0.073	0.064	0.075	0.063	0.076	0.058	0.074
0.1	1,000	0.238	0.238	0.210	0.211	0.196	0.199	0.196	0.199	0.179	0.183
	2,000	0.414	0.408	0.371	0.364	0.350	0.354	0.309	0.328	0.316	0.320
0.15	1,000	0.885	0.854	0.847	0.818	0.818	0.793	0.784	0.762	0.768	0.746
	2,000	0.974	0.956	0.966	0.951	0.956	0.933	0.946	0.925	0.941	0.919
0.2	1,000	0.976	0.924	0.972	0.916	0.964	0.893	0.961	0.889	0.956	0.887
	2,000	0.992	0.951	0.989	0.945	0.987	0.923	0.987	0.914	0.985	0.910

Third, the power of $\Pi(l)$ and $\Pi_v(l)$ is comparable, but the former needs a longer computational time. Note that when $\nu_0 = (15, 10)$ and $(20, 10)$, the test results are similar to those for $\nu_0 = (10, 8)$, and hence are not reported for brevity.

Overall, both estimators $\widehat{\theta}$ and $\widehat{\theta}_v$ and both tests $\Pi(l)$ and $\Pi_v(l)$ exhibit good performance, especially when the sample size T gets larger. When the dimension of Y_t is small, our simulation results show that $\widehat{\theta}_v$ is only slightly less efficient than $\widehat{\theta}$, and $\Pi_v(l)$ is, in general, as powerful as $\Pi(l)$. When the dimension of Y_t is large, $\widehat{\theta}_v$ and $\Pi_v(l)$ enjoy faster computation speeds than those of $\widehat{\theta}$ and $\Pi(l)$, respectively. As such, we recommend using $\widehat{\theta}_v$ and $\Pi_v(l)$ in practice.

7. Applications

In this section, we consider two applications to the U.S. stock market. Application 1 studies the low-dimensional RCOV matrix series calculated using the composite realized kernels (CRK) in Lunde, Shephard and Sheppard (2016). Application 2 studies the high-dimensional RCOV series calculated using the TARVM estimator in Tao et al. (2011).

7.1. Application 1

In this application, we revisit the RCOV matrix data of Hewlett-Packard Development Company, L.P. (HPQ), International Business Machines Corporation (IBM), and Microsoft Corporation (MSFT) in Lunde, Shephard and Sheppard (2016). This data set, denoted by $\{Y_t\}_{t=1}^{1474}$, ranges from January 2006 to December 2011, with 1,474 observations in total. Here, two flash crashes are flagged on May 6, 2010, and August 9, 2011, and are replaced by an average of the nearest

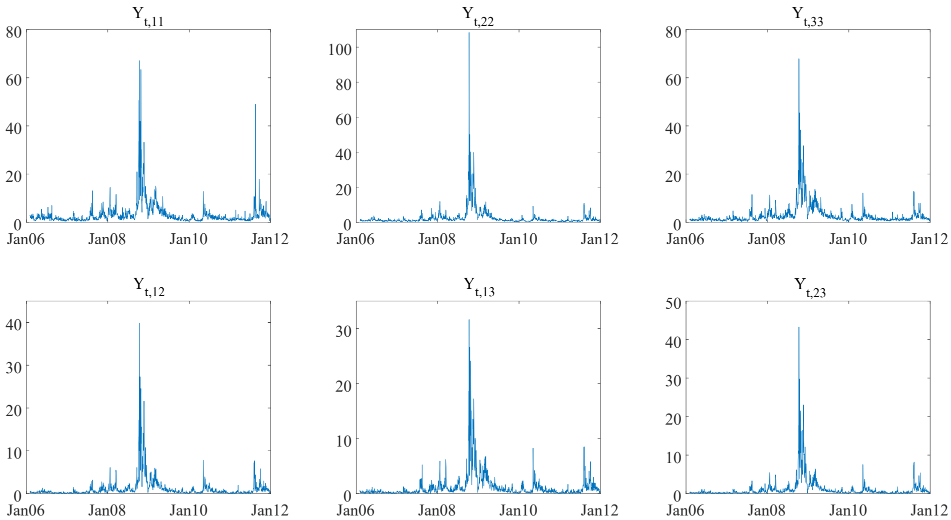

Figure 2. Components of Y_t.

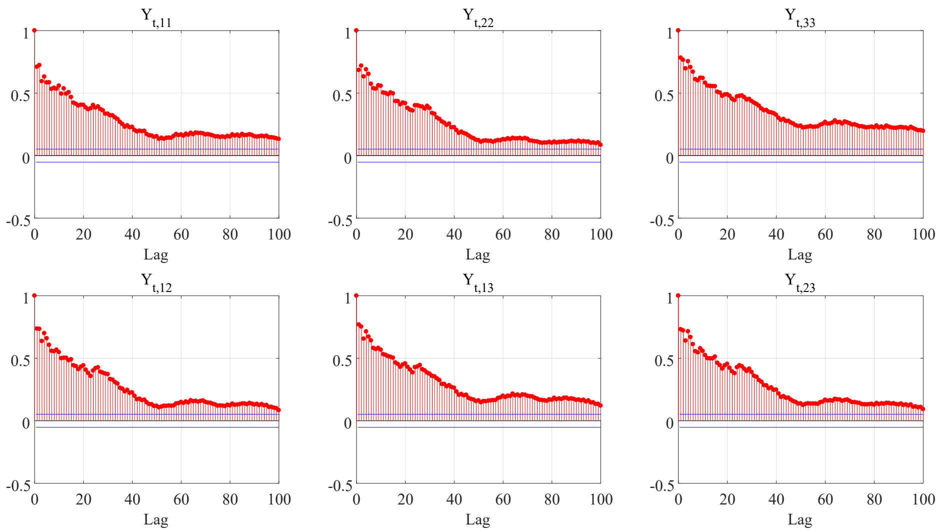

Figure 3. Sample ACFs of each component $Y_{t,ij}$.

five preceding and following matrices.

Figure 2 plots the diagonal and off-diagonal components of $\{Y_t\}_{t=1}^{1474}$, showing that Y_t has a clear clustering feature. Figure 3 plots their sample autocorrelation functions (ACFs), which show the significant temporal dependence of Y_t.

Table 3. The results of the estimated diagonal VT-CBF and VT-CBF-HAR models.

Diagonal VT-CBF model								
$\widehat{\nu}_v$		$\widehat{S}_v$		$\widehat{A}_{11,v}$	$\widehat{B}_{11,v}$	$\widehat{B}_{12,v}$	$\widehat{B}_{13,v}$	persistence
74.0110	3.1523	1.1099	1.1635	0.7207	0.5358	0.0117	0.4129	0.9771
(10.7545)	(1.8844)	(0.9031)	(0.7705)	(0.0223)	(0.0365)	(0.0176)	(0.0354)	
40.5849	1.1099	2.3683	1.0965	0.7200	0.5620	0.0119	0.3800	0.9788
(3.9787)	(0.9031)	(2.1165)	(0.9209)	(0.0246)	(0.0289)	(0.0177)	(0.0382)	
	1.1635	1.0965	2.7883	0.7118	0.5579	0.0127	0.3977	0.9762
	(0.7705)	(0.9209)	(1.3276)	(0.0211)	(0.0292)	(0.0190)	(0.0354)	
Diagonal VT-CBF-HAR model								
$\widehat{\nu}_v$		$\widehat{S}_v$		$\widehat{A}_{(d),v}$	$\widehat{A}_{(w),v}$	$\widehat{A}_{(m),v}$	persistence	
69.0222	3.1523	1.1099	1.1635	0.6954	0.5735	0.3891	0.9639	
(6.2261)	(2.2543)	(1.0464)	(0.8881)	(0.0256)	(0.0443)	(0.0344)		
40.4021	1.1099	2.3683	1.0965	0.6884	0.6027	0.3557	0.9637	
(2.9408)	(1.0464)	(2.3391)	(1.0210)	(0.0275)	(0.0318)	(0.0426)		
	1.1635	1.0965	2.7883	0.6703	0.6041	0.3812	0.9596	
	(0.8881)	(1.0210)	(1.4971)	(0.0279)	(0.0318)	(0.0364)		

† The asymptotic standard errors are given in parentheses.

Based on these facts, we first fit $\{Y_t\}_{t=1}^{1474}$ using a diagonal VT-CBF model with $(P, Q, K) = (1, 3, 1)$, where the order K is taken as one for ease of model identification, and the orders P and Q are selected using the Bayesian information criterion (BIC). Specifically, this diagonal VT-CBF model is estimated using the two-step estimation procedure, and the corresponding estimates are give in Table 3. Second, because the sample ACFs of each component in Figure 3 decay slowly, we also fit $\{Y_t\}_{t=1}^{1474}$ using a diagonal VT-CBF-HAR model and list the related estimation results in Table 3. From this table, we find that the estimates of the degrees of freedom (especially for ν_2) in both fitted models are close to each other, and both estimates of ν_2 are small, indicating the heavy-tailedness of the examined data. For the estimates of the mean parameter matrix S, its standard errors based on the VT-CBF model are smaller than those based on the VT-CBF-HAR model. For other estimates of the parameter matrices, the estimated diagonal components in each parameter matrix seem to have similar values, meaning that the three stocks possibly have similar temporal structures. This similarity can also be seen from the persistence values of each stock in Table 3, where the persistence of stock s is defined by $\sum_{i=1}^{P} A_{1i,ss}^2 + \sum_{j=1}^{Q} B_{1j,ss}^2$ for the VT-CBF model and $A_{(d),ss}^2 + A_{(w),ss}^2 + A_{(m),ss}^2$ for the VT-CBF-HAR model. After the estimation, we apply our test statistics $\Pi_v(l)$ to both fitted models, and the results summarized in Table 4 imply that both fitted models are adequate at the 5% level.

Table 4. The results of $\Pi_v(l)$ for the diagonal VT-CBF and VT-CBF-HAR models.

	Diagonal VT-CBF model					Diagonal VT-CBF-HAR model				
l	2	3	4	5	6	2	3	4	5	6
$\Pi_v(l)$	1.494	4.170	8.004	9.428	11.513	4.385	6.127	7.004	10.310	11.583
p-value	0.474	0.244	0.091	0.093	0.074	0.112	0.106	0.136	0.067	0.072

Next, we consider the forecasting performance of our proposed diagonal VT-CBF and VT-CBF-HAR models. Specifically, we compute the one-step, five-step and ten-step predictions of the RCOV matrices based on a rolling window procedure, with the window size equal to $T_0 = 800$. That is, for $T_0 \leq t \leq T - t_0$, we fit the models based on T_0 observations $\{Y_s\}_{s=t-T_0+1}^{t}$, and forecast $\widehat{Y}_{t+t_0}$ with $t_0 = 1, 5, 10$, and calculate the forecasting error as $\widehat{Y}_{t+t_0} - Y_{t+t_0}$. To examine the importance of ν_2 in the CBF models, we also apply the diagonal VT-CAW and VT-CAW-HAR models to perform predictions. The diagonal VT-CAW and VT-CAW-HAR models are defined in the same way as the diagonal VT-CAW and VT-CAW-HAR models, except that the matrix-F distribution for Δ_t in the latter two models is replaced by the Wishart distribution. In addition to the CAW-type models, we further include a diagonal VAR-HAR model for comparison, where this VAR model uses a HAR structure with the diagonal autoregressive parameter matrices to fit $y_t = vech(Y_t)$.

Table 5 gives the average forecasting errors in Frobenius and spectral norms for all models. Here, we find that, regardless of the prediction horizon, the diagonal VT-CBF-HAR model always has the smallest forecasting error in both norms. Moreover, we apply the DM test of Diebold and Mariano (1995) to examine whether the diagonal VT-CBF-HAR model has significantly better forecasting accuracy than those of the other four competing models. The corresponding results are given in Table 5, and show that the VT-CBF-HAR model is significantly better than its four competing models in terms of the five-step and ten-step forecasts. For one-step forecasts, the VT-CBF-HAR and VT-CBF model models have comparable forecasting accuracy, and the VT-CBF-HAR model is significantly better than the remaining three models at the 10% level. Note that the VAR-HAR model always performs worst in all examined cases, probably because it disentangles the matrix-structure of the RCOV matrices, which may have some intrinsic and useful value for forecasts.

Table 5. Forecasting errors based on different models and the related DM testing results.

	1-step		5-step		10-step	
Diagonal Model	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral
VT-CBF-HAR	1.5284	1.4607	1.9725	1.8850	2.2108	2.1091
VT-CBF	1.5349	1.4664	1.9955	$1.9069^{\dagger}$	2.2802^{*}	2.1755^{*}
VT-CAW-HAR	1.5383^{*}	1.4703^{*}	2.0029^{*}	1.9147^{*}	$2.2864^{\diamond}$	$2.1813^{\diamond}$
VT-CAW	1.5390	1.4699	$2.0253^{\diamond}$	$1.9351^{\diamond}$	$2.3364^{\diamond}$	$2.2286^{\diamond}$
VAR-HAR	$1.6472^{\diamond}$	$1.5661^{\diamond}$	$2.1700^{\diamond}$	$2.0626^{\diamond}$	$2.6088^{\diamond}$	$2.4711^{\diamond}$

The DM test is used to compare the prediction accuracy between the diagonal VT-CBF-HAR and the other four competing models. The result for each competing model is marked with a "$\dagger$", "$*$" or "$\diamond$" if the DM test implies that the Diagonal VT-CBF-HAR model gives significantly more accurate predictions than this competing model at the 10%, 5%, or 1% level, respectively.

7.2. Application 2

In this section, we consider intraday data of 112 stocks from four major sectors constituting the S&P 500 index: 31 stocks from the financial sector, 31 stocks from the industrial sector, 25 stocks from the health care sector, and 25 stocks from consumer discretionary sector; see the full lists of stocks in Appendix S4. All intraday price data are downloaded from the Wharton Research Data Services (WRDS) database, and are taken from July 1, 2009, to December 30, 2016, including a total of 1,890 non-missing dates of trading data. Based on 100 times log of the price data, the daily RCOV matrices $\{Y_t\}_{t=1}^{1890}$ are calculated using the TARVM method of Tao et al. (2011) for each sector.

For each sector, because the dimension of the RCOV matrix is large, we fit the RCOV matrix data using the diagonal F-VT-CBF and F-VT-CBF-HAR models. To do this, we first look for the value of r in model (5.11) by plotting the ratios $\{\lambda_i/\lambda_{i+1}\}$ for each sector in Figure 4, where $\{\lambda_i\}$ are the eigenvalues of $\bar{S}$ in descending order. From Figure 4, we can choose $r = 3$ for the financial sector, $r = 2$ for the industrial sector, $r = 2$ for the health care sector, and $r = 1$ for the consumer discretionary sector. To get more information, we also plot the ratios $\{\lambda_i/\lambda_{i+1}\}$ for all four pooled sectors in Figure 5, which suggests $r = 3$. This implies that all 112 stocks considered may be driven by three latent factors. However, only two may affect the industrial and health care sectors, and only one may affect the consumer discretionary sector. Hence, it is more reasonable to study the RCOV matrix data across sectors rather than together.

Next, we estimate the diagonal F-VT-CBF and F-VT-CBF-HAR models, and choose the orders using a similar procedure as in Application 1. The related results are reported in Table 6. From this table, we find that except for the mean parameter matrix, the diagonal components of other parameter matrices

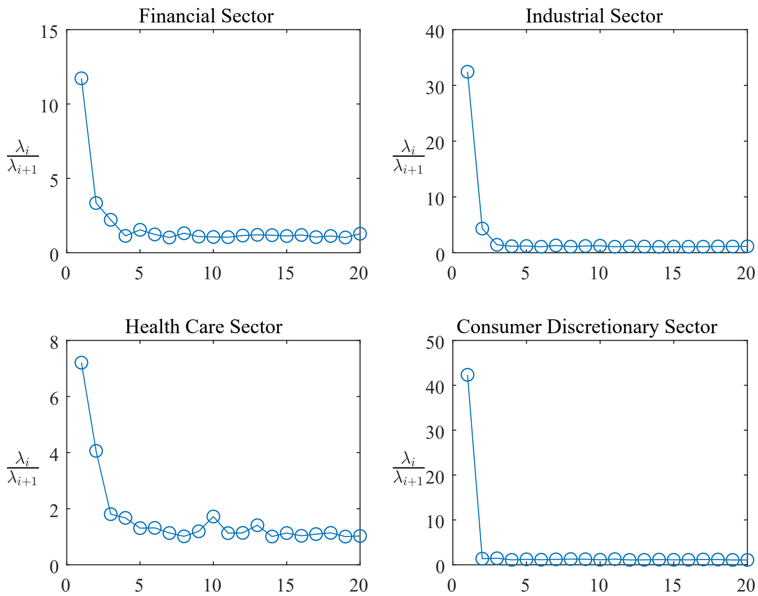

Figure 4. Ratios of adjacent eigenvalues of $\bar{S}$ for each sector.

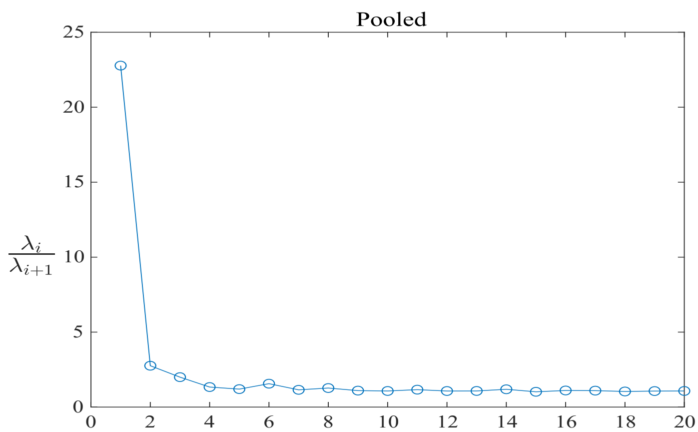

Figure 5. Ratios of adjacent eigenvalues of $\bar{S}$ for all four pooled sectors.

seem to have different values, meaning that each component of Y_{ft} has a different dynamical structure. Moreover, the values of persistence for $Y_{ft,ss}$ show clear differences across the four sectors, with the largest persistence in the financial sector and the smallest persistence in the health care sector. This finding indicates that the effect of past stock returns to its current volatility decays very slowly in

Table 6. The results of the estimated diagonal F-VT-CBF and F-VT-CBF-HAR models.

Diagonal F-VT-CBF model										
Sector	$\widehat{\nu}_{fv}$		$\widehat{S}_{fv}$		$\widehat{A}_{11,fv}$	$\widehat{B}_{11,fv}$	$\widehat{B}_{12,fv}$	$\widehat{B}_{13,fv}$	$\widehat{B}_{14,fv}$	persistence
Financial	35.3380	25.7553	0.6808	0.1389	0.7269	0.5118	0.2741	0.3219		0.9691
	(2.9679)	(11.0314)	(2.3577)	(0.6519)	(0.0348)	(0.0518)	(0.1014)	(0.0606)		
	19.257	0.6808	2.5799	0.0211	0.6844	0.5382	0.3172	0.3628		0.9903
	(1.0419)	(2.3577)	(9.6931)	(0.1730)	(0.0608)	(0.1181)	(0.1831)	(0.0699)		
		0.1389	0.0211	1.6309	0.7292	0.3010	0.4490	0.3817		0.9696
		(0.6519)	(0.1730)	(1.8857)	(0.0732)	(0.1468)	(0.0897)	(0.1201)		
Industrial	24.9287	17.3161	2.1513		0.7277	0.6488				0.9505
	(6.9460)	(7.0877)	(1.0290)		(0.0729)	(0.0709)				
	22.7808	2.1513	1.0614		0.6716	0.6921				0.9300
	(7.6622)	(1.0290)	(0.3786)		(0.0317)	(0.0373)				
Health Care	24.3415	8.6744	3.4402		0.7617	0.5396	0.1129			0.8841
	(4.9720)	(2.9442)	(0.7505)		(0.1324)	(0.0651)	(0.6685)			
	15.9965	3.4402	2.185		0.7351	0.5706	0.0001			0.8660
	(5.1757)	(0.7505)	(0.4998)		(0.1407)	(0.1585)	(0.8598)			
Consumer Discretionary	22.4570	15.3282			0.7516	0.4517	0.2604	0.1971	0.2666	0.9467
	(4.0371)	(4.9315)			(0.0261)	(0.0724)	(0.1171)	(0.1711)	(0.1032)	
	12.2757									
	(1.4843)									

Diagonal F-VT-CBF-HAR model								
Sector	$\widehat{\nu}_{fv}$		$\widehat{S}_{fv}$		$\widehat{A}_{(d),fv}$	$\widehat{A}_{(w),fv}$	$\widehat{A}_{(m),fv}$	persistence
Financial	38.0409	25.7553	0.6808	0.1389	0.7041	0.5069	0.4573	0.9618
	(3.1046)	(15.5296)	(2.6814)	(0.8796)	(0.0259)	(0.0830)	(0.1098)	
	18.9242	0.6808	2.5799	0.0211	0.6676	0.4588	0.5739	0.9855
	(0.8746)	(2.6814)	(10.6104)	(0.2816)	(0.0441)	(0.1162)	(0.0628)	
		0.1389	0.0211	1.6309	0.7659	0.2502	0.5476	0.9491
		(0.8796)	(0.2816)	(1.2904)	(0.0484)	(0.1678)	(0.0537)	
Industrial	25.0002	17.3161	2.1513		0.7161	0.5494	0.3549	0.9406
	(5.9220)	(10.0000)	(1.2538)		(0.0699)	(0.0758)	(0.0458)	
	22.3305	2.1513	1.0614		0.6361	0.6086	0.3283	0.8830
	(6.7511)	(1.2538)	(0.4310)		(0.0462)	(0.0970)	(0.1484)	
Health Care	23.3766	8.6744	3.4402		0.7259	0.5357	0.1944	0.8625
	(3.6648)	(3.2870)	(0.8134)		(0.1095)	(0.1141)	(0.0369)	
	16.1320	3.4402	2.1850		0.6961	0.5689	0.0691	0.8130
	(4.6804)	(0.8134)	(0.5280)		(0.0918)	(0.1620)	(0.2421)	
Consumer Discretionary	23.1216	15.3282			0.7285	0.4865	0.4092	0.9348
	(3.2789)	(6.0954)			(0.0299)	(0.0599)	(0.0502)	
	11.9375							
	(1.1630)							

† The asymptotic standard errors given in parentheses are based on $\widehat{Y}_{ft}$ rather than Y_{ft}.

the financial sector, but behaves oppositely in the health care sector.

Finally, we examine the forecasting performance of our F-CBF models. As in Application 1, five different diagonal factor models (see Table 7) are considered to forecast Y_t, based on a rolling window procedure with a window size equal to 1,000. Their forecasting performance is evaluated using the average of the forecasting errors in the Frobenius and spectral norms as well as the results of the related DM test in Table 7. From this table, we can see that except for the health care sector, the diagonal F-VT-CBF-HAR model always has the smallest forecasting error and the diagonal F-VAR-HAR model has the largest forecasting error. For one-step forecasts in the health care sector, the diagonal F-VT-CAW-HAR has a slightly smaller forecasting error than that of the diagonal F-VT-CBF-HAR model. In view of the results of the DM test, the diagonal F-VT-CBF-HAR model exhibits a significantly better performance than the other four models in terms of five-step and ten-step forecasts. However, this advantage is slightly weak in terms of one-step forecasts, for which the diagonal F-VT-CBF and F-VT-CAW-HAR models have similar performance in the industrial sector, and the diagonal F-VT-CAW-HAR and F-VAR-HAR models have comparative performance in the health care sector.

8. Conclusion

We propose a new CBF model to study the dynamics of RCOV matrices. For this CBF model, we explore its stationarity and moment properties, establish the asymptotics of its MLE, and investigate inner-product-based tests for its model checking. Hence, a systematic inferential tool for this CBF model is available for empirical researchers. In order to deal with large-dimensional RCOV matrices, we also construct two reduced CBF models: the VT-CBF model and the F-CBF model. For both reduced models, the asymptotic theory of the estimated parameters is derived. Compared with the CAW model with Wishart innovations, the CBF model with matrix-F innovations is better able to capture the heavy-tailed RCOV. This advantage is demonstrated by two real examples on U.S. stock markets. As motivated by Chiriac and Voev (2011), an obvious future work is to introduce a fractional integration structure into our CBF models. Furthermore, we can extend the idea of using the matrix-F innovation in a number of ways, resulting in a large family of models. This is important in terms of studying the positive definite dynamics.

Supplementary Material

The online Supplementary Material contains the proofs of all theorems, some useful derivatives, and the stock lists used in the second application.

Table 7. Forecasting errors based on different factor models and the related DM testing results.

Sector	Diagonal Model	1-step		5-step		10-step	
		Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral
Financial	F-VT-CBF-HAR	8.7701	7.9339	10.4581	9.7229	11.0221	10.3200
	F-VT-CBF	8.8116	7.9824†	10.6677*	9.9315◇	11.3503*	10.6713◇
	F-VT-CAW-HAR	8.7865	7.9644*	10.5183	9.8144†	11.1072	10.4575
	F-VT-CAW	8.8354*	8.0248◇	10.7097*	10.0151*	11.5030◇	10.8786◇
	F-VAR-HAR	8.8878*	8.0662*	11.1055◇	10.4644◇	11.7725◇	11.1745◇
Industrial	F-VT-CBF-HAR	7.9567	7.0936	9.3154	8.5480	9.8270	9.0842
	F-VT-CBF	7.9735	7.1169	9.4094	8.6334	9.9837	9.2397
	F-VT-CAW-HAR	7.9680	7.1112†	9.4106*	8.6494*	10.0565◇	9.3255◇
	F-VT-CAW	7.9995*	7.1450*	9.4645*	8.7001*	10.1157*	9.3826*
	F-VAR-HAR	8.0567*	7.2170*	9.6801◇	8.9531◇	10.2809◇	9.5794◇
Health Care	F-VT-CBF-HAR	6.6253	5.8586	7.4977	6.8076	7.8436	7.1863
	F-VT-CBF	6.6628†	5.9019†	7.6400*	6.9605*	8.0708◇	7.4398◇
	F-VT-CAW-HAR	6.6126	5.8559	7.5658*	6.8892*	7.9743◇	7.3317◇
	F-VT-CAW	6.7451◇	6.0117◇	8.0423◇	7.3944◇	8.3738◇	7.7569◇
	F-VAR-HAR	6.6688	5.8954	7.6163*	6.9389*	7.9457†	7.2872
Consumer Discretionary	F-VT-CBF-HAR	8.3355	7.0130	9.3278	8.1225	9.6830	8.5081
	F-VT-CBF	8.3552†	7.0415*	9.4191†	8.2195†	9.8426*	8.6883*
	F-VT-CAW-HAR	8.3517*	7.0307*	9.3886◇	8.1935◇	9.7918◇	8.6294◇
	F-VT-CAW	8.3727*	7.0560◇	9.4489*	8.2546*	9.9211◇	8.7754◇
	F-VAR-HAR	8.3914*	7.0762*	9.5017*	8.3282*	9.9085◇	8.7575◇

The DM test is used to compare the prediction accuracy between the diagonal F-VT-CBF-HAR and the other four competing models. The result for each competing model is marked with a "†", "$*$" or "$\diamond$" if the DM test implies that the Diagonal F-VT-CBF-HAR model gives significantly more accurate predictions than this competing model at the 10%, 5%, or 1% level, respectively.

Acknowledgments

The authors greatly appreciate the helpful comments and suggestions of the two anonymous referees, associate editor, and co-editor. Jiang was supported by the China Scholarship Council (No.201906210093) and NSFC (Nos.11571348 and 11771239), and acknowledges that some of the work was carried out during the visits to the University of Hong Kong and University of Illinois at Urbana-Champaign, and at Tsinghua University. Zhu's work was supported in part by the Hong Kong GRF grant Nos. 17304421, 17306818, 17305619, NSFC (Nos. 11690014, 11731015 and 71532013), Seed Fund for Basic Research (No. 201811159049), and Fundamental Research Funds for the Central University (19JNYH08). Li's work was supported in part by the Hong Kong GRF grant (No. 17303315) and EdUHK grant RG44/2019–2020R.

References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81**, 1203–1227.

Aït-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics* **201**, 384–399.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P (2003). Modeling and forecasting realized volatility. *Econometrica* **71**, 579–625.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 253–280.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* **72**, 885–925.

Barndorff-Nielsen, O. E., Hansen, P. R. and Lunde, A. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* **162**, 149–169.

Bauer, G. H. and Vorkink, K. (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics* **160**, 93–101.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* **69**, 542–547.

Bollerslev, T., Engle, R. F. and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy* **96**, 116–131.

Bollerslev, T., Patton, A. J. and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics* **192**, 1–18.

Boussama, F., Fuchs, F. and Stelzer, R. (2011). Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and their Applications* **121**, 2331–2360.

Callot, L., Kock, A. and Medeiros, M. (2017). Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics* **32**, 140–158.

Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics* **26**, 922–947.

Comte, F. and Lieberman, O. (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* **84**, 61–84.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* **7**, 174–196.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* **13**, 253–263.

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* **20**, 339–350.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* **11**, 122–150.

Engle, R. F. and Lee, G. (1999). A long-run and short-run component model of stock return volatility. In *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive WJ Granger* (Edited by R. F. Engle and H. White), 475–497. Oxford University Press, Oxford.

Engle, R. F. and Mezrich, J. (1996). GARCH for groups. *Risk* **9**, 36–40.

Fan, J., Qi, L. and Xiu, D. (2014). Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business & Economic Statistics* **32**, 178–191.

Francq, C., Horváth, L. and Zakoïan, J.-M. (2011). Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics* **9**, 619–656.

Golosnoy, V., Gribisch, B. and Liesenfeld, R. (2012). The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics* **167**, 211–223.

Gouriéroux, C., Jasiak, J. and Sufana, R. (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics* **150**, 167–181.

Hafner, C. M. and Preminger, A. (2009). On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis* **100**, 2044–2054.

Hansen, P. R., Huang, Z. and Shek, H. H. (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* **27**, 877–906.

Jin, X. and Maheu, J. M. (2013). Modeling realized covariances and returns. *Journal of Financial Econometrics* **11**, 335–369.

Jin, X. and Maheu, J. M. (2016). Bayesian semiparametric modeling of realized covariance matrices. *Journal of Econometrics* **192**, 19–39.

Kim, D., Kong, X., Li, C. and Wang, Y. (2018). Adaptive thresholding for large volatility matrix estimation based on high-frequency financial data. *Journal of Econometrics* **203**, 69–79.

Konno, Y. (1991). A note on estimating eigenvalues of scale matrix of the multivariate F-distribution. *Annals of the Institute of Statistical Mathematics* **43**, 157–165.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *The Annals of Statistics* **40**, 694–726.

Leung, P. L. and Lo, M. (1996). An identity for the noncentral multivariate F distribution with application. *Statistica Sinica* **6**, 419–431.

Li, D., Zhang, X., Zhu, K. and Ling, S. (2018). The ZD-GARCH model: A new way to study heteroscedasticity. *Journal of Econometrics* **202**, 1–17.

Li, W. K. and McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **43**, 231–239.

Ling, S. and Li, W. K. (1997). Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors. *Journal of Time Series Analysis* **18**, 447–464.

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 Statistic* **34**, 504–518.

McAleer, M. and Medeiros, M. C. (2008). Realized volatility: A review. *Econometric Reviews* **27**, 10–45.

McCurdy, T. H. and Stengos, T. (1992). A comparison of risk-premium forecasts implied by parametric versus nonparametric conditional mean estimators. *Journal of Econometrics* **52**, 225–244.

Noureldin, D., Shephard, N. and Sheppard, K. (2012). Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* **27**, 907–933.

Oh, D. H. and Patton, A. J. (2017). Modeling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics* **35**, 139–154.

Opschoor, A., Janus, P., Lucas, A. and Dijk, D. (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics* **36**, 643-657.

Pedersen, R. S. and Rahbek, A. (2014). Multivariate variance targeting in the BEKK-GARCH model. *Econometrics Journal* **17**, 24–55.

Tao, M., Wang, Y., Yao, Q. and Zou, J.(2011). Large volatility matrix inference via combining low-frequency and high-frequency approaches. *Journal of the American Statistical Association* **106**, 1025–1040.

Tao, M., Wang, Y. and Zhou, H. H. (2013). Optimal sparse volatility matrix estimation for high-dimensional Itô processes with measurement errors. *The Annals of Statistics* **41**, 1816–1864.

Tse, Y. K. (2002). Residual-based diagnostics for conditional heteroscedasticity models. *Econometrics Journal* **5**, 358–374.

Wang, Y. and Zou, J.(2010). Vast volatility matrix estimation for high-frequency financial data. *The Annals of Statistics* **38**, 943–978.

Yu, P. L., Li, W. K. and Ng, F. C.(2017). The generalized conditional autoregressive Wishart model for multivariate realized volatility. *Journal of Business & Economic Statistics* **35**, 513–527.

Zhu, K. and Li, W. K. (2015). A new pearson-type QMLE for conditionally heteroscedastic models. *Journal of Business & Economic Statisticss* **33**, 552–565.

Jiayuan Zhou

Department of Statistics, University of Florida, Gainesville, FL 32611, USA.

E-mail: zhou.j@ufl.edu

Feiyu Jiang

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

E-mail: jiangfy@fudan.edu.cn

Ke Zhu

Department of Statistics & Actuarial Science, The University of Hong Kong, Hong Kong, China.

E-mail: mazhuke@hku.hk

Wai Keung Li

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

E-mail: hrntlwk@hku.hk

(Received November 2019; accepted August 2020)