The Annals of Statistics
2024, Vol. 52, No. 3, 1027–1049
https://doi.org/10.1214/24-AOS2381

HIGH-DIMENSIONAL COVARIANCE MATRICES UNDER DYNAMIC VOLATILITY MODELS: ASYMPTOTICS AND SHRINKAGE ESTIMATION

BY YI DING[1,a] AND XINGHUA ZHENG[2,b]

[1]*Faculty of Business Administration, University of Macau,* [a]*yiding@um.edu.mo*
[2]*Department of ISOM, Hong Kong University of Science and Technology,* [b]*xhzheng@ust.hk*

We study the estimation of high-dimensional covariance matrices and their empirical spectral distributions under dynamic volatility models. Data under such models have nonlinear dependency both cross-sectionally and temporally. We establish the condition under which the limiting spectral distribution (LSD) of the sample covariance matrix under scalar BEKK models is different from the i.i.d. case. We then propose a time-variation adjusted (TV-adj) sample covariance matrix and prove that its LSD follows the Marčenko–Pastur law. Based on the asymptotics of the TV-adj sample covariance matrix, we develop a consistent population spectrum estimator and an asymptotically optimal nonlinear shrinkage estimator of the unconditional covariance matrix.

1. Introduction.

1.1. *The Marčenko–Pastur law.* Random matrix theory (RMT) is a powerful tool in the study of high-dimensional statistics. When the dimension and sample size grow proportionally, for i.i.d. data, it is well known that the limiting spectral distribution (LSD) of the sample covariance matrix is connected to that of the population covariance matrix through the Marčenko–Pastur equation; see, for example, Marčenko and Pastur (1967), Yin (1986), Silverstein (1995), and Silverstein and Bai (1995). El Karoui (2008) studies estimating population spectrum based on the Marčenko–Pastur (M-P) law, and Ledoit and Wolf (2012, 2015, 2020) develop algorithms for estimating the population spectrum and nonlinear shrinkage estimation of the covariance matrix. All these studies focus on the case where the observations are i.i.d.

1.2. *Dynamic volatility models.* An important feature of financial returns is that their volatilities are time-varying and dependent over time. Dynamic volatility models such as the multivariate generalized autoregressive conditional heteroskedasticity (GARCH) (Engle, Granger and Kraft (1984), Bollerslev, Engle and Wooldridge (1988)), the dynamic conditional correlation (DCC) model (Engle (2002)), and the Baba, Engle, Kraft, and Kroner (BEKK) model (Engle and Kroner (1995)) are popular in studying the dynamic variances and covariances. In particular, the widely used scalar BEKK model (Ding and Engle (2001)) describes the dynamics of the covariance matrix as follows:

$$\mathbf{\Sigma}_{t+1} = (1-a-b)\overline{\mathbf{\Sigma}} + a\mathbf{R}_t\mathbf{R}_t^T + b\mathbf{\Sigma}_t, \tag{1.1}$$

where $\mathbf{\Sigma}_t$ is the conditional covariance matrix, $\overline{\mathbf{\Sigma}}$ is the unconditional covariance matrix of the returns $\mathbf{R}_t = (R_{1t}, \dots, R_{pt})^T$, and $0 < a, b < 1$ with $a+b < 1$ are the related parameters. The parameter a is sometimes referred to as the innovation coefficient, and $a+b$ the persistence coefficient.

Received December 2023.
MSC2020 subject classifications. Primary 62B20; secondary 62H12, 62M10.
Key words and phrases. High-dimension, dynamic volatility model, sample covariance matrix, spectral distribution, nonlinear shrinkage.

To estimate a dynamic volatility model, a common approach is variance/correlation targeting (Pedersen and Rahbek (2014), Pakel et al. (2021)). The method requires estimating the unconditional covariance/correlation matrix. When the dimension is high, the sample covariance/correlation matrix is not consistent. For large dynamic volatility models, estimating the unconditional covariance/correlation matrix is challenging and calls for rigorous investigation.

1.3. *Existing research in RMT for sample covariance matrix when there is time dependency.* There is a growing literature on the study of limiting spectral properties of the sample covariance matrix when there is time dependency. Jin et al. (2009), Yao (2012), Liu, Aue and Paul (2015), and Bhattacharjee and Bose (2016) obtain the LSD of the sample covariance/autocovariance matrix of linearly dependent time series that can be transformed into data with independent columns. Banna and Merlevède (2015) and Merlevède and Peligrad (2016) investigate the LSD of the sample covariance matrix of stationary dependent processes with independent rows. Yaskov (2017) focuses on the case where data have dependence in finite lags. Zheng and Li (2011) establish the LSD, and Yang, Zheng and Chen (2021) derive the central limit theorem of linear spectral statistics of sample covariance matrix under elliptical models.

Engle, Ledoit and Wolf (2019) propose to estimate the unconditional covariance/correlation matrix under large BEKK/DCC models using the nonlinear shrinkage (NLS) estimator developed in Ledoit and Wolf (2012, 2015). Ledoit and Wolf (2012, 2015) document that the NLS estimator has several advantages in estimating the high-dimensional covariance matrix. For example, it does not rely on sparsity assumptions on the covariance matrix, and for i.i.d. data, it is consistent in estimating the asymptotically optimal shrinkage estimator in the class of rotation-equivariant estimators; see Ledoit and Wolf (2012, 2015) for detailed explanations. It is worth emphasizing that the asymptotic property of the NLS estimator relies on the fact that the LSD of the sample covariance matrix follows the M-P law.

1.4. *Our contributions.* We aim to estimate the unconditional covariance matrix under large dynamic volatility models. An important and natural question motivated by the proposal of Engle, Ledoit and Wolf (2019) is: *Does the NLS estimator work under large dynamic volatility models?*

To see how the dynamic volatility model can affect the spectral distribution of the sample covariance matrix, we simulate data from BEKK model (1.1) with $\overline{\mathbf{\Sigma}} = \mathbf{I}$, $a = 0.05$, and $b = 0.9$, which is the setting used in Engle, Ledoit and Wolf (2019). The dimension $p = 100$ or 500, and the sample size n satisfies $p/n = 0.8$. We compute the empirical spectral distribution (ESD) of the sample covariance matrix and compare it with the M-P distribution. The results are shown in Figure 1. We see from Figure 1 that the ESD of the sample covariance matrix under the BEKK model substantially deviates from the M-P law. Therefore, it is problematic to perform NLS on the sample covariance matrix the same way as in the i.i.d. case.

In this paper, we investigate the limiting spectral properties of the sample covariance matrix under large BEKK models. We show that if $\eta(a, b, p) := (a/(1 - a - b)) \min(\sqrt{p(1 - a - b)}, 1) \to 0$, then the LSD of the sample covariance matrix shares the same limit as the i.i.d. case; see Theorem 1. We call this case the reducible case. On the other hand, if $\eta(a, b, p)$ is bounded away from zero, then the spectral distribution of the sample covariance under the BEKK model is more heavy-tailed than the i.i.d. case; see Theorem 2. We call this case the non-reducible case. Under the non-reducible case, reversing the M-P law can not consistently estimate the population spectrum, and the NLS estimator of the covariance matrix is not asymptotically optimal.

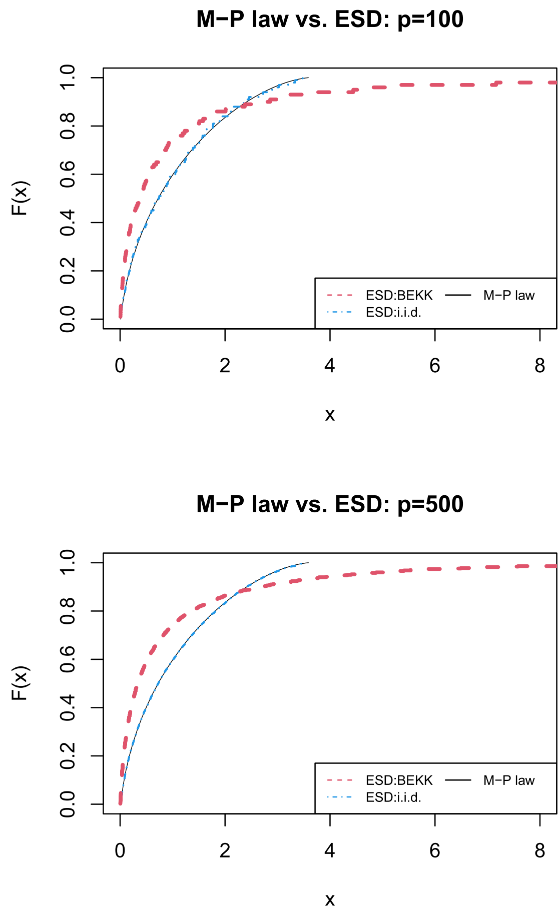

FIG. 1. *Empirical spectral distributions of sample covariance matrices under the BEKK model, compared with the ESD based on i.i.d. data and the M-P distribution.*

Next, we address the problem of population spectrum estimation under large BEKK models. We first estimate the parameters a and b using a quasi-maximum likelihood estimator (QMLE) from univariate GARCH models. Next, we develop a time-variation tracking matrix, $\mathbf{P}_t$, which can track the time-variation in the conditional covariance matrices. We then define time-variation adjusted returns, $\widetilde{\mathbf{R}}_t = \mathbf{P}_t^{-1/2}\mathbf{R}_t$, and a *time-variation adjusted (TV-adj) sample covariance matrix*, $\widetilde{\mathbf{S}}_n = \sum_{t=1}^{n} \widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t^T/n$. We prove that the TV-adj sample covariance matrix shares the same LSD as the i.i.d. case; see Theorem 3. Using the TV-adj sample covariance matrix and existing M-P law reversing algorithms, we obtain a TV-adj shrinkage estimator of the population spectrum, which we show is consistent; see Corollary 1 for the exact statement.

Finally, we tackle the problem of unconditional covariance matrix estimation. We develop a TV-adj nonlinear shrinkage (TV-Adj NLS) estimator and show that it consistently estimates the asymptotically optimal shrinkage estimator; see Theorem 5.

In summary, our contributions lie in the following aspects. First, we establish the condition under which the LSD of the sample covariance matrix under large BEKK models is different from the i.i.d. case. Second, we propose a TV-adj sample covariance matrix and develop an estimator that can consistently recover the population spectral distribution. Third, we develop a TV-adj NLS estimator and prove that it is asymptotically optimal.

The rest of this paper is organized as follows. The main theoretical results are given in Section 2. Simulation studies are presented in Section 3. We conclude in Section 4. The proof of Theorem 3 is presented in Section 5. The proofs of other main results and additional simulation results are collected in the Supplementary Material (Ding and Zheng (2024)).

The following notation is used throughout the paper. For any matrix $\mathbf{A} = (A_{ij})$, its spectral norm is defined as $\|\mathbf{A}\| = \max_{\|\mathbf{x}\| \le 1} \sqrt{\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}}$, where $\|\mathbf{x}\| = \sqrt{\sum x_i^2}$ for any vector $\mathbf{x} = (x_i)$; the Frobenius norm is defined as $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} A_{ij}^2}$. We write $\mathbf{A} \ge 0 (> 0)$ if the matrix $\mathbf{A}$ is positive semi-definite (positive definite), and $\mathbf{A} \ge \mathbf{B} (> \mathbf{B})$ if $\mathbf{A} - \mathbf{B} \ge 0 (> 0)$. If $\mathbf{A} \ge 0$, $\mathbf{A}^{1/2}$ is defined as the positive semi-definite matrix that satisfies $(\mathbf{A}^{1/2})^2 = \mathbf{A}$. For any symmetric matrix $\mathbf{A}$ with eigenvalues $\lambda_1, \dots, \lambda_p$, its empirical spectral distribution (ESD) is defined as $F^{\mathbf{A}}(x) = \sum_{j=1}^p \mathbf{1}_{[\lambda_j, +\infty)}(x)/p$, $x \in \mathbb{R}$. For two sequences of positive real numbers (a_n) and (b_n), we write $a_n \gg b_n$ if $a_n/b_n \to \infty$, and $a_n \ll b_n$ if $a_n/b_n \to 0$. Finally, we use $\xrightarrow{\text{P}}$ to represent convergence in probability.

2. Main results.

2.1. *Setting and assumptions.* Under a dynamic volatility model, returns are modeled as $\mathbf{R}_t = (\mathbf{\Sigma}_t)^{1/2}\mathbf{z}_t$, where $\mathbf{z}_t = (z_{1t}, \dots, z_{pt})^T$ are i.i.d. with mean zero and covariance matrix $\mathbf{I}$. We suppose that $(\mathbf{R}_t)$ follows the scalar BEKK model (1.1). Define $\mathbf{R}_t^0 = (\overline{\mathbf{\Sigma}})^{1/2}\mathbf{z}_t$, $t = 1, \dots, n$, which share the same unconditional covariance matrix as $\mathbf{R}_t$ but are i.i.d. Denote the corresponding sample covariance matrices as follows:

$$\mathbf{S}_n = \frac{1}{n}\sum_{t=1}^n \mathbf{R}_t\mathbf{R}_t^T \quad \text{and} \quad \mathbf{S}_n^0 = \frac{1}{n}\sum_{t=1}^n \mathbf{R}_t^0(\mathbf{R}_t^0)^T.$$

We write $\widehat{\lambda}_1 \ge \cdots \ge \widehat{\lambda}_p$ as the eigenvalues of $\mathbf{S}_n$, and $\widehat{\lambda}_1^0 \ge \cdots \ge \widehat{\lambda}_p^0$ as the eigenvalues of $\mathbf{S}_n^0$.

We impose the following assumptions.

ASSUMPTION 1.

(i) $\mathbf{z}_t \underset{\text{i.i.d.}}{\sim} N(0, \mathbf{I})$.
(ii) $\overline{\mathbf{\Sigma}}$ is nonnegative definite and its ESD, $F^{\overline{\mathbf{\Sigma}}}$, converges in distribution to a probability distribution H on $[0, \infty)$ as $p \to \infty$, and $H \ne \delta(0)$, the Dirac measure at 0.
(iii) $\|\overline{\mathbf{\Sigma}}\| < C$ for some constant $C > 0$.
(iv) The dimension p and the sample size n satisfy that $p, n \to \infty$, and $p/n \to y > 0$.

About the parameters a and b, we allow them to depend on p. Specifically, we denote by a_p and b_p the coefficients in the BEKK model when the dimension is p.

2.2. *Limiting properties of ESD of sample covariance matrix under large BEKK model.*

2.2.1. *Reducible case.* Note that if $a_p = 0$, then the BEKK model reduces to the i.i.d. case with $\mathbf{\Sigma}_t \equiv \overline{\mathbf{\Sigma}}$. In general, if a_p is close to 0, then the BEKK model will be similar to the i.i.d. case.

Recall that for any two distributions, F_1 and F_2, the Levy distance between them is defined as

$$L(F_1, F_2) := \inf\{\varepsilon > 0 | F_1(x - \varepsilon) - \varepsilon \le F_2(x) \le F_1(x + \varepsilon) + \varepsilon \text{ for all } x \in \mathbb{R}\}.$$

It is well known that convergence in Levy distance implies convergence in distribution.

Define

$$\eta(a_p, b_p, p) = \frac{a_p}{1 - a_p - b_p}\min\left(\sqrt{p(1 - a_p - b_p)}, 1\right). \tag{2.1}$$

The next theorem shows that $F^{\mathbf{S}_n}$ and $F^{\mathbf{S}_n^0}$ are asymptotically indistinguishable when $\eta(a_p, b_p, p) \to 0$.

THEOREM 1. *Under model* (1.1) *and Assumption* 1, *if* $\eta(a_p, b_p, p) \to 0$ *as* $p \to \infty$, *then*

$$L\left(F^{\mathbf{S}_n}, F^{\mathbf{S}_n^0}\right) = o_p(1). \tag{2.2}$$

Theorem 1 implies that when $\eta(a_p, b_p, p) \to 0$, the ESD of sample covariance matrix under the BEKK model converges to the same M-P law as the i.i.d. case. For example, when $1 - a_b - b_p$ is bounded away from zero, namely, the model is not near-integration, then $a_p \to 0$ implies that the M-P law is retained. On the other hand, in the near-integration case when $1 - a_b - b_p \to 0$, to retain the M-P law, a_p needs to converge to zero at a faster rate than $\min(\sqrt{p/(1 - a_p - b_p)}, 1/(1 - a_p - b_p))$.

Below, we refer to the case when $\eta(a_p, b_p, p) \to 0$ as the *reducible case*. Under the reducible case, the population spectrum can be recovered by reversing the M-P law, and the NLS estimator of the covariance matrix is asymptotically optimal.

2.2.2. *Non-reducible case.* When the reducible condition does not hold, what will the ESD of the sample covariance matrix be like? We have seen in Figure 1 that when $a = 0.05$ and $b = 0.9$, which is a typical setting calibrated from empirical data (Engle, Ledoit and Wolf (2019)), the ESD under the BEKK model appears to be more heavy-tailed than the i.i.d. case. We refer to the case when $\eta(a_p, b_p, p)$ is bounded away from zero as the *non-reducible case*. In practice, the two coefficients a_p and b_p learned from financial data appear to fit the non-reducible case. Therefore, investigating the non-reducible case is not only of theoretical interest but also practically relevant. It can be easily shown that $E(\mathrm{tr}(\mathbf{S}_n)) = E(\mathrm{tr}(\mathbf{S}_n^0))$. We compare the ESD's under the BEKK model and the i.i.d. case by their second moments:

$$M_2 = M_2^p = \frac{1}{p}\sum_{i=1}^{p}\widehat{\lambda}_i^2 = \frac{1}{p}\mathrm{tr}\left((\mathbf{S}_n)^2\right) \quad \text{and} \quad M_2^0 = M_2^{0,p} = \frac{1}{p}\sum_{i=1}^{p}(\widehat{\lambda}_i^0)^2 = \frac{1}{p}\mathrm{tr}\left((\mathbf{S}_n^0)^2\right).$$

For the i.i.d. case, $E(M_2^{0,p}) = yH_1^2 + H_2 + o(1)$, where $H_1 = \lim_{p\to\infty}\mathrm{tr}(\overline{\boldsymbol{\Sigma}})/p$ and $H_2 = \lim_{p\to\infty}\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^2)/p$; see equation (4.14) of Yin (1986). The next theorem states that $E(M_2^p) > E(M_2^{0,p})$ when $\eta(a_p, b_p, p)$ is bounded away from zero.

THEOREM 2. *Under model* (1.1) *and Assumption* 1, *if* $\eta(a_p, b_p, p) > c$ *for some constant* $c > 0$, *then there exists* $\delta > 0$ *such that for all* p *large enough*,

$$E\left(M_2^p\right) \geq E\left(M_2^{0,p}\right) + \delta. \tag{2.3}$$

2.3. *Time-variation adjusted spectrum estimator.* Theorems 1 and 2 suggest that, the usual spectrum estimator based on the M-P law does not always work under the BEKK model. To recover the population spectrum under the BEKK model, new estimators need to be developed.

Zheng and Li (2011) study a similar problem under elliptical models. They propose a self-normalization approach to remove the time-variation in the covariance matrices. Specifically, under the elliptical model, $\boldsymbol{\Sigma}_t = \xi_t\boldsymbol{\Sigma}_0$, where (ξ_t) is a one-dimensional process. The process (ξ_t) can be consistently estimated by $(\|\mathbf{R}_t\|^2/p)$ as the dimension $p \to \infty$. Therefore, normalizing $\mathbf{R}_t$ by $\|\mathbf{R}_t\|$ removes the time-variation in the covariance matrix of $\mathbf{R}_t$. Motivated by this idea, we aim to adjust the dynamic volatilities of the nonlinearly dependent data so that the adjusted data behave asymptotically i.i.d. Under the BEKK model, the time-variation is more complicated because the cross-sectional dependence is dynamic and nonlinearly dependent on past returns. Removing the time-variation under the BEKK model relies on an innovative way to reverse transform the observations.

Our approach is as follows. We first estimate the parameters a_p and b_p. Under the BEKK model (1.1), each (R_{it}) follows a univariate GARCH model:

$$\sigma_{i,t+1}^2 = (1 - a_p - b_p)\overline{\sigma}_i^2 + a_p R_{it}^2 + b_p \sigma_{i,t}^2, \tag{2.4}$$

where $\sigma_{i,t}^2 = (\mathbf{\Sigma}_t)_{ii}$, and $\overline{\sigma}_i^2 = (\overline{\mathbf{\Sigma}})_{ii}$ for $1 \le i \le p$. As a result, a_p and b_p can be estimated without knowing the whole unconditional covariance matrix. Specifically, we randomly select one variable, say, i_0, fit a univariate GARCH model to $(R_{i_0 t})$ and get QMLE $\widehat{a}_p$ and $\widehat{b}_p$:

$$(\widehat{a}_p, \widehat{b}_p, \widehat{\overline{\sigma}}_{i_0}) = \underset{(a,b,\overline{\sigma}_{i_0})\in\Omega}{\operatorname{argmax}} -\sum_{t=1}^{n}\left(\frac{R_{i_0 t}^2}{\sigma_{i_0,t}^2} + \log(\sigma_{i_0,t}^2)\right),$$

where $\Omega = \{(a, b, \overline{\sigma}) : 0 \le a, b \le a + b < 1, \delta \le \overline{\sigma} < C\}$, $\delta, C > 0$ are constants. The QMLE of the univariate GARCH model is consistent with a convergence rate of $\sqrt{n}$ when $a_p + b_p$ is bounded away from one; see, for example, Theorems 2.1 and 2.2 of Francq and Zakoïan (2004). Below, we give the convergence result for the case when a_p is close to zero and b_p is close to one.

PROPOSITION 1. *Suppose that there is a sequence of GARCH processes, $(R_t) = (R_{t;p})$, which satisfy $R_t = \sigma_t z_t$, $(z_t)'s$ are i.i.d., $E(z_t^2) = 1$, $E(z_t^K) < \infty$ for all $K \ge 1$, $\sigma_{t+1}^2 = (1 - a_p - b_p)\sigma^2 + a_p R_t^2 + b_p \sigma_t^2$, and $0 < a_p, b_p < a_p + b_p < 1$. Suppose in addition that a_p, b_p, and the sample size n satisfy that $a_p \asymp 1 - a_p - b_p \to 0$ as $p \to \infty$, and $(1 - a_p - b_p) \gg n^{-\upsilon}$ for some $\upsilon \in (0, 1/4)$. Then there exists a local QMLE that satisfies, for all p large enough,*

$$\widehat{a}_p - a_p = O_p\left(\frac{1}{n^{1/2-\varepsilon}}\right) \quad \text{and} \quad \widehat{b}_p - b_p = O_p\left(\frac{1}{n^{1/2-\varepsilon}}\right), \tag{2.5}$$

where ε is any positive constant.

We then use $\widehat{a}_p$, $\widehat{b}_p$ and past returns to construct a time-variation tracking matrix:

$$\mathbf{P}_t = \max\left(\frac{1 - \widehat{a}_p - \widehat{b}_p + \widehat{a}_p \widehat{b}_p^{K_p}}{1 - \widehat{b}_p}, \kappa\right)\mathbf{I} + \sum_{j=1}^{K_p} \widehat{a}_p \widehat{b}_p^{j-1} \mathbf{R}_{t-j} \mathbf{R}_{t-j}^T, \tag{2.6}$$

where K_p represents the number of lagged returns, which grows with p at a rate specified in Theorem 3 below, and κ is a small positive constant. The intuition behind such a definition is that, $\mathbf{\Sigma}_t = (1 - a_p - b_p)/(1 - b_p)\overline{\mathbf{\Sigma}} + \sum_{j=1}^{\infty} a_p b_p^{j-1} \mathbf{R}_{t-j} \mathbf{R}_{t-j}^T$, hence an appropriate choice of K_p will make $\mathbf{P}_t^{-1/2} \mathbf{\Sigma}_t \mathbf{P}_t^{-1/2}$ close to $\overline{\mathbf{\Sigma}}$; see equation (5.24) for the precise statement.

We use $\mathbf{P}_t^{-1/2}$ to reverse transform the observed returns. Specifically, we define the time-variation adjusted returns:

$$\widetilde{\mathbf{R}}_t = \mathbf{P}_t^{-1/2} \mathbf{R}_t. \tag{2.7}$$

We will call $\mathbf{P}_t^{-1/2}$ the reverse transformation matrix. Using $\widetilde{\mathbf{R}}_t$, we construct the time-variation adjusted (TV-adj) sample covariance matrix:

$$\widetilde{\mathbf{S}}_n = \frac{1}{n} \sum_{t=1}^{n} \widetilde{\mathbf{R}}_t (\widetilde{\mathbf{R}}_t)^T. \tag{2.8}$$

THEOREM 3. *Under model* (1.1) *and Assumption* 1, *suppose that either one of the following conditions holds*:

(i) $a_p p^{1/2} \to \infty$, $b_p p^{\varpi} \to \infty$, $1 - a_p - b_p > \delta > 0$ *for some constants* $\delta > 0$ *and* $\varpi \in (0, 1/2)$, *and* K_p *satisfies that* $K_p \to \infty$ *and* $K_p \ll p^{1/2-\varpi}$;

(ii) $a_p \asymp 1 - a_p - b_p \to 0$ *as* $p \to \infty$, $1 - a_p - b_p \gg n^{-\upsilon}$ *for some* $\upsilon \in (0, 1/4)$, *and* K_p *satisfies that* $p^{\upsilon} \log p \ll K_p \ll p^{1/2-\varepsilon}$ *for some* $\varepsilon > 0$.

Then

$$L\big(F^{\widetilde{\mathbf{S}}_n}, F^{\mathbf{S}_n^0}\big) = o_p(1). \tag{2.9}$$

REMARK 1. The proof of Theorem 3 relies on the rotational invariance of the distribution of $\mathbf{z}_t$. This is the reason that we assume $(\mathbf{z}_t)$ follows a standard multivariate normal distribution. In the RMT literature, observations are typically assumed to be $\boldsymbol{\Sigma}^{1/2}\mathbf{z}_t$, where $\mathbf{z}_t$ contains independent standardized entries. In this formulation, no rotational transformation is allowed, namely, the observations can not be $\boldsymbol{\Sigma}^{1/2}\mathcal{O}_t\mathbf{z}_t$, where $(\mathcal{O}_t)$ is a sequence of orthogonal matrices. If the M-P law can be proved for observations given by $\boldsymbol{\Sigma}^{1/2}\mathcal{O}_t\mathbf{z}_t$, then the normality assumption can be removed in our setting. Numerical studies seem to support this; see Appendix B.1 of the supplement material (Ding and Zheng (2024)) for more details. We will leave the question of whether the normality assumption can be removed for future research.

Theorem 3 implies that the time-variation adjusted sample covariance matrix has the same LSD as the i.i.d. case. In particular, $F^{\widetilde{\mathbf{S}}_n} \xrightarrow{\mathrm{P}} F$, where F is determined by H in that its Stieltjes transform,

$$m_F(z) := \int_{\lambda\in\mathbb{R}} \frac{1}{\lambda - z}\, dF(\lambda), \quad z \in \mathbb{C}^+ := \big\{z \in \mathbb{C}, \operatorname{Im}(z) > 0\big\}, \tag{2.10}$$

solves the following equation:

$$m_F(z) = \int_{\tau\in\mathbb{R}} \frac{1}{\tau(1 - y(1 + z m_F(z))) - z}\, dH(\tau); \tag{2.11}$$

see, for example, Theorem 1 in Marčenko and Pastur (1967).

We can then consistently estimate the population spectrum by reversing the M-P equation. Specifically, we denote the eigenvalues of $\widetilde{\mathbf{S}}_n$ by $\widetilde{\lambda}_1 \ge \cdots \ge \widetilde{\lambda}_p$. We first regularize the eigenvalues of $\widetilde{\mathbf{S}}_n$ to be $\widetilde{\lambda}_i^{\tau} = \min(\widetilde{\lambda}_i, L)$ for some large constant L. We then apply the Quantized Eigenvalues Sampling Transform (QuEST) algorithm in Ledoit and Wolf (2015) to $(\widetilde{\lambda}_i^{\tau})'s$ and obtain the estimated population spectrum. Denote by $\widehat{\lambda}_1^H \ge \widehat{\lambda}_2^H \ge \cdots \ge \widehat{\lambda}_p^H$ the estimated eigenvalues and $\lambda_1^H \ge \lambda_2^H \ge \cdots \ge \lambda_p^H$ the eigenvalues of $\overline{\boldsymbol{\Sigma}}$.

COROLLARY 1. *Under the assumptions of Theorem* 3, *if in addition* $y \ne 1$, *then*

$$\frac{1}{p}\sum_{i=1}^{p}\big(\lambda_i^H - \widehat{\lambda}_i^H\big)^2 = o_p(1).$$

Corollary 1 guarantees that QuEST applied to the TV-adj sample covariance matrix consistently estimates the population spectrum of the unconditional covariance matrix.

2.4. *Time-variation adjusted nonlinear shrinkage estimator of unconditional covariance matrix*. The NLS estimator (Ledoit and Wolf (2012, 2015, 2020)) is consistent in estimating the asymptotically optimal shrinkage estimator for i.i.d. data. In financial applications, the NLS has gained popularity for large portfolio optimization; see, for example, Ledoit and Wolf (2017), Ao, Li and Zheng (2019), De Nard, Ledoit and Wolf (2021) and Ding, Li and Zheng (2021).

Motivated by the NLS developed under the i.i.d. case, to estimate the unconditional covariance matrix under large BEKK models, we use the TV-adj sample covariance matrix and consider rotation-equivariant shrinkage estimators in the form $\widehat{\mathbf{\Sigma}} = \sum_{i=1}^{p} \widehat{d}_i \widetilde{u}_i \widetilde{u}_i^T$, where $(\widetilde{u}_i)_{1\le i\le p}$ are eigenvectors of the TV-adj sample covariance matrix $\widetilde{\mathbf{S}}_n$. The optimal rotation-equivariant estimator finds $(\widehat{d}_i)_{1\le i\le p}$ that minimize $\|\widehat{\mathbf{\Sigma}} - \overline{\mathbf{\Sigma}}\|_F$. Elementary algebra shows that the optimal solution is $\widehat{d}_i^* = \widetilde{u}_i^T \overline{\mathbf{\Sigma}} \widetilde{u}_i$.

In search of the asymptotically optimal shrinkage formula under large BEKK models, we study the following generalized empirical spectral distribution of the TV-adj sample covariance matrix:

$$F^{\widetilde{\mathbf{S}}_n, g(\overline{\mathbf{\Sigma}})}(x) = \frac{1}{\operatorname{tr}(g(\overline{\mathbf{\Sigma}}))} \sum_{i=1}^{p} \big(\widetilde{u}_i^T g(\overline{\mathbf{\Sigma}}) \widetilde{u}_i\big) \cdot \mathbf{1}_{[\widetilde{\lambda}_i, +\infty)}(x), \tag{2.12}$$

which generalizes the ESD of $\widetilde{\mathbf{S}}_n$ by replacing the weight $1/p$ with $\widetilde{u}_i^T g(\overline{\mathbf{\Sigma}})\widetilde{u}_i / \operatorname{tr}(g(\overline{\mathbf{\Sigma}}))$ for some bounded function $g(\cdot)$, and $g(\overline{\mathbf{\Sigma}}) = \sum_{i=1}^{p} g(\lambda_i^H) v_i v_i^T$, where v_i's are the eigenvectors of $\overline{\mathbf{\Sigma}}$. The limit of the generalized ESD of the sample covariance matrix under the i.i.d. case is obtained in Ledoit and Péché (2011) and is used to derive the asymptotically optimal shrinkage estimator. Parallel to the i.i.d. case, we study the limiting property of the generalized ESD of the TV-adj sample covariance matrix via the following generalized Stieltjes transform:

$$\Theta_n^g(z) = \frac{1}{p} \operatorname{tr}\big((\widetilde{\mathbf{S}}_n - z\mathbf{I})^{-1} g(\overline{\mathbf{\Sigma}})\big). \tag{2.13}$$

The following theorem gives the limit of $\Theta_n^g(z)$.

THEOREM 4. *Under the assumptions of Theorem 3, if, in addition, the limiting distribution H is supported by $[h_1, h_2]$ for some constants $0 < h_1 \le h_2 < \infty$, and g is a bounded function on $[h_1, h_2]$ with finitely many points of discontinuity, then*

$$\Theta_n^g(z) - \Theta^g(z) = o_p(1) \quad \textit{for all } z \in \mathbb{C}^+,$$

where

$$\Theta^g(z) = \int_{-\infty}^{+\infty} \big(\tau\big(1 - y^{-1} - y^{-1} z m_F(z)\big) - z\big)^{-1} g(\tau)\, dH(\tau). \tag{2.14}$$

The function $\Theta^g(z)$ is the limit of the generalized Stieltjes transform under the i.i.d. case; see Theorem 2 of Ledoit and Péché (2011). Theorem 4 states that the generalized ESD based on the time-variation adjusted sample covariance matrix converges to the same limit as the i.i.d. case. Therefore, we can utilize the same nonlinear shrinkage algorithm that is developed for i.i.d. case to obtain the time-variation adjusted nonlinear shrinkage estimator under BEKK models.

Specifically, to estimate the unconditional covariance matrix, we perform the nonlinear shrinkage algorithm by Ledoit and Wolf (2015) on $\widetilde{\mathbf{S}}_n^\tau$, where $\widetilde{\mathbf{S}}_n^\tau = \sum_{i=1}^{p} \widetilde{\lambda}_i^\tau \widetilde{u}_i \widetilde{u}_i^T$, $\widetilde{\lambda}_i^\tau = \min(\widetilde{\lambda}_i, L)$, and L is a large constant. The truncation is applied to ensure that the support of the ESD is bounded. We denote by $\widetilde{\boldsymbol{\Sigma}}$ the resulting covariance matrix estimator, which we call the time-variation adjusted nonlinear shrinkage (TV-adj NLS) estimator. Define $\widetilde{\boldsymbol{\Sigma}}^{\mathrm{or}} = \sum_{i=1}^{p} d_i^{\mathrm{or}}(\widetilde{\lambda}_i^\tau)\widetilde{u}_i\widetilde{u}_i^T$, where

$$(2.15)\quad d_i^{\mathrm{or}}(\widetilde{\lambda}_i^\tau) = \begin{cases} \dfrac{1}{(y-1)\check{m}_{\underline{F}}(0)} & \text{if } \widetilde{\lambda}_i^\tau = 0 \text{ and } y > 1, \\ \dfrac{\widetilde{\lambda}_i^\tau}{(1 - y - y\widetilde{\lambda}_i^\tau \cdot \check{m}_F(\widetilde{\lambda}_i^\tau))^2} & \text{otherwise}, \end{cases} \quad \text{for } i = 1, \dots, p,$$

$\underline{F}(x) = (1-y)\mathbf{1}_{\{[0,\infty)\}}(x) + yF(x)$, $m_{\underline{F}}(z) = (y-1)/z + ym_F(z)$, $\check{m}_F(\lambda) = \lim_{z \in \mathbb{C}^+ \to \lambda} m_F(z)$, and $F(\cdot)$ and $m_F(z)$ are given in equations (2.10) and (2.11), respectively. By Theorem 4 and Theorem 4 of Ledoit and Péché (2011), $\widetilde{\boldsymbol{\Sigma}}^{\mathrm{or}}$ is the infeasible oracle shrinkage estimator.

THEOREM 5. *Under the assumptions of Theorem 4, if in addition $y \neq 1$, then*

$$\frac{1}{\sqrt{p}}\|\widetilde{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{\Sigma}}^{\mathrm{or}}\|_F = o_p(1).$$

Theorem 5 guarantees that the TV-adj NLS consistently estimates the oracle shrinkage estimator under dimension-normalized Frobenius norm. The convergence rate achieved by the TV-adj NLS under BEKK models matches with that of the ordinary NLS under the i.i.d. case; see Proposition 4.3 of Ledoit and Wolf (2012) and Theorem 3.1 of Ledoit and Wolf (2015).

3. Simulation studies.

3.1. *Simulation setup.* We generate data from the BEKK model (1.1) with $\mathbf{z}_t \underset{\text{i.i.d.}}{\sim} N(0, \mathbf{I})$. The unconditional covariance matrix is set to be $\overline{\boldsymbol{\Sigma}} = (\rho^{|i-j|})_{1 \le i,j \le p}$, where $\rho = 0.4$.[1] The dimension is set to be $p = 100$ or 500. We fix $p/n = 0.8$. Additional simulation results for different p, n ratios and other simulation settings are collected in Appendix B of the supplement material (Ding and Zheng (2024)).

About the parameters (a, b), first, we choose four cases: $(a, b) \in \{(0, 0), (0.15, 0.25), (0.1, 0.65), (0.05, 0.9)\}$. The setting $(a, b) = (0, 0)$ corresponds to the i.i.d. case, which is presented as a benchmark, and the other (a, b) pairs correspond to nontrivial BEKK cases sorted with increasing magnitudes of $\eta(a, b, p)$ defined in equation (2.1), representing increasing levels of deviation from the i.i.d. case. The last configuration $(a, b) = (0.05, 0.9)$ is the setting used in Engle, Ledoit and Wolf (2019). We simulate 100 replications under each setting and present the results in Section 3.2.1.

Next, we examine more choices of (a, b). Specifically, we consider a grid of (a, b)'s in the region $\{(a, b) : 0.05 \le a \le 0.5, 0.05 \le b \le 0.90, a + b \le 0.95\}$ and show the results from 100 replications in Section 3.2.2.

3.2. *Simulation results.*

3.2.1. *Four (a, b) cases.* In this subsection, we present the simulation results for four (a, b) cases: $(a, b) \in \{(0, 0), (0.15, 0.25), (0.1, 0.65), (0.05, 0.9)\}$.

[1] The results for the settings where $\rho = 0, 0.2, 0.6$ and 0.8 are qualitatively similar.

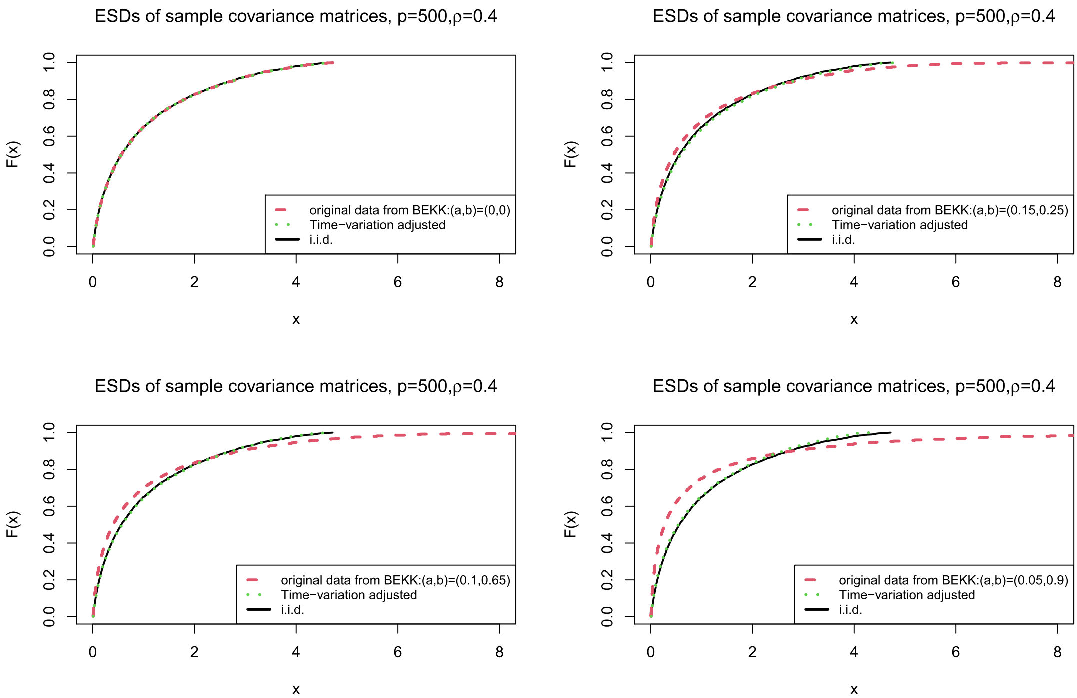

FIG. 2. *ESDs of sample covariance matrices of the original data, the time-variation adjusted data, and the i.i.d. data for* $p = 500$, $n = 625$. *The unconditional covariance matrix* $\overline{\boldsymbol{\Sigma}} = (0.4)^{|i-j|}$; $(a, b) \in \{(0, 0), (0.15, 0.25), (0.1, 0.65), (0.05, 0.9)\}$.

Empirical spectral distribution of the sample covariance matrices. We compute the original sample covariance matrix $\mathbf{S}_n = \sum_{t=1}^n \mathbf{R}_t\mathbf{R}_t^T/n$ and the TV-adj sample covariance matrix $\widetilde{\mathbf{S}}_n = \sum_{t=1}^n \widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t^T/n$, and compare their ESDs with that of the sample covariance matrix under the i.i.d. case, namely, $\mathbf{S}_n^0 = \sum_{t=1}^n \mathbf{R}_t^0(\mathbf{R}_t^0)^T/n$.

We first illustrate the ESDs from one random realization for $p = 500$ in Figure 2. We see that for all four cases, the ESDs of the TV-adj sample covariance matrix match remarkably well with that of the i.i.d. case. On the other hand, under nontrivial BEKK models, the ESDs of the original sample covariance matrices deviate from the M-P law, in particular, they are more heavy-tailed.

We then perform 100 replications and summarize the Euclidean distance between the eigenvalues of $\mathbf{S}_n^0$ and $\widetilde{\mathbf{S}}_n$ or $\mathbf{S}_n$ in Table 1. We see from Table 1 that under nontrivial BEKK settings, the distance between the ESD of the original sample covariance matrix and that under the i.i.d. case increases with increasing magnitudes of $\eta(a, b, p)$. The distance between the ESD of the TV-adj sample covariance matrix and that under the i.i.d. case is smaller and closer to zero under various nontrivial BEKK settings. Moreover, it decreases as p gets larger.

Population spectrum estimation. Next, we evaluate the estimators of the population eigenvalues. We compare the performance of the proposed time-variation adjusted NLS spectrum estimator[2] (TV-adj NLS-Spectrum) with that of the NLS spectrum estimator based on the original sample covariance matrix (original NLS-Spectrum). We measure the estimation error by $\sqrt{\sum_{1\le i\le p}(\lambda_i^H - \widehat{\lambda}_i^H)^2}$, where $\widehat{\lambda}_1^H \ge \widehat{\lambda}_2^H \ge \cdots \ge \widehat{\lambda}_p^H$ are the estimated eigenvalues, and $\lambda_1^H \ge \lambda_2^H \ge \cdots \ge \lambda_p^H$ are the eigenvalues of $\overline{\boldsymbol{\Sigma}}$.

In Figure 3, we plot the distributions of the estimated eigenvalues from one random realization with $p = 500$. We see that, under all four cases, the proposed TV-adj spectrum

[2]The function "tau_estimate" from R package "nlshrink" is used to compute the estimated eigenvalues.

TABLE 1

Summary of the distance $\sqrt{\sum_{1\le i\le p}(\widehat{\lambda}_i - \widehat{\lambda}_i^0)^2}$, *where* $(\widehat{\lambda}_i^0)_{1\le i\le p}$ *are eigenvalues of* $\mathbf{S}_n^0$, *and* $(\widehat{\lambda}_i)_{1\le i\le p}$ *are eigenvalues of* $\mathbf{S}_n$ *or* $\widetilde{\mathbf{S}}_n$. *The table shows the mean and standard deviation* (*in parenthesis*) *from* 100 *replications*

(a, b)		(0.15, 0.25)	(0.1, 0.65)	(0.05, 0.9)
$(p, n) = (100, 125)$	$\mathbf{S}_n$	0.277	0.413	0.907
		(0.078)	(0.106)	(0.195)
	$\widetilde{\mathbf{S}}_n$	0.089	0.123	0.215
		(0.034)	(0.033)	(0.079)
$(p, n) = (500, 625)$	$\mathbf{S}_n$	0.279	0.429	1.046
		(0.045)	(0.050)	(0.120)
	$\widetilde{\mathbf{S}}_n$	0.029	0.054	0.162
		(0.017)	(0.029)	(0.056)

estimator is close to the population spectrum, and its performance is similar to the spectrum estimator based on the i.i.d. data. On the other hand, the shrinkage spectrum estimator based on the original sample covariance matrix significantly deviates from the population spectrum. Table 2 shows the Euclidean distances between the estimated population eigenvalues and the true ones from 100 replications. We see from Table 2 that the error of the original NLS-spectrum estimator increases with $\eta(a, b, p)$. It also gets larger as the dimension gets higher. The proposed TV-adj NLS-spectrum estimator dominantly outperforms the original NLS-spectrum estimator with a substantially lower estimation error. The performance of the TV-adj NLS-spectrum estimator is only slightly worse than the infeasible shrinkage estimator based on i.i.d. data for the first two nontrivial BEKK settings. For the fourth setting when $(a, b) = (0.05, 0.9)$, because $a + b$ is close to one, the estimation error of the TV-adj NLS-

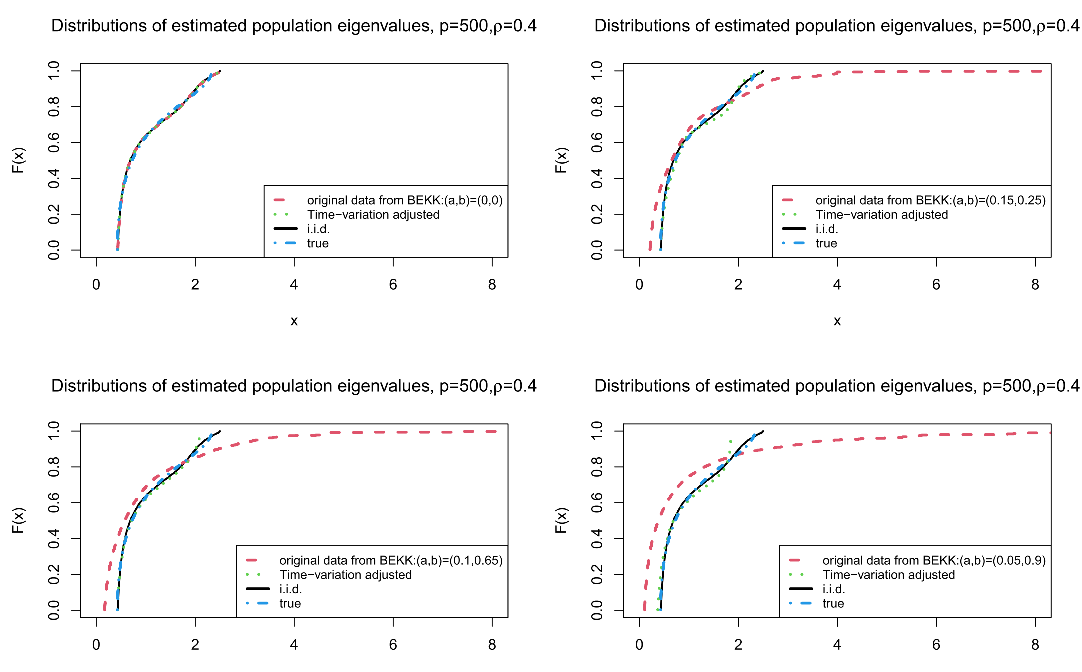

FIG. 3. *Distributions of estimated eigenvalues. They are obtained by nonlinear shrinkage estimators applied to the original data, the time-variation adjusted data, and the i.i.d. data. The dimension* $p = 500$, $n = 625$. $\overline{\mathbf{\Sigma}} = (0.4)^{|i-j|}$; $(a, b) \in \{(0, 0), (0.15, 0.25), (0.1, 0.65), (0.05, 0.9)\}$.

TABLE 2
Summary of the distance between the estimated eigenvalues and the population ones. We report the mean and standard deviation (in parenthesis) from 100 replications. The pair $(a, b) = (0, 0)$ represents the i.i.d. case and is presented as the benchmark. The remaining (a, b) pairs are for nontrivial BEKK cases

		i.i.d.	BEKK		
	(a, b)	$(0, 0)$	$(0.15, 0.25)$	$(0.1, 0.65)$	$(0.05, 0.9)$
$(p, n) = (100, 125)$	original NLS-Spectrum	0.136	0.411	0.566	1.109
		(0.041)	(0.094)	(0.124)	(0.209)
	TV-adj NLS-Spectrum	0.143	0.217	0.251	0.313
		(0.047)	(0.071)	(0.071)	(0.100)
$(p, n) = (500, 625)$	original NLS-Spectrum	0.052	0.401	0.577	1.243
		(0.026)	(0.049)	(0.053)	(0.119)
	TV-adj NLS-Spectrum	0.054	0.057	0.088	0.184
		(0.027)	(0.025)	(0.026)	(0.035)

spectrum estimator is larger. However, as the dimension p grows, the error decreases and becomes closer to that of the shrinkage estimator under the i.i.d. case.

Unconditional covariance matrix estimation. Finally, we evaluate the unconditional covariance matrix estimation. We compute the NLS estimators[3] based on the time-variation adjusted sample covariance matrix (TV-adj NLS) and the original sample covariance matrix (original NLS). The estimation error is measured by the Frobenius norm $\sqrt{\sum_{1\le i,j\le p}(\overline{\boldsymbol{\Sigma}}_{ij} - \widehat{\boldsymbol{\Sigma}}_{ij})^2}$, where $\widehat{\boldsymbol{\Sigma}}$ is the estimated unconditional covariance matrix. The results are summarized in Table 3. We see from Table 3 that the estimation error of the original NLS increases sharply as $\eta(a, b, p)$ gets large and as the dimension grows. The proposed TV-adj NLS greatly improves over the original NLS with a substantially lower estimation error. The performance of the TV-adj NLS is only slightly worse than that of the NLS under the i.i.d. case for the first two nontrivial (a, b) settings. For the most challenging case

TABLE 3
Summary statistics of the estimation error of the estimated unconditional covariance matrix in Frobenius norm $\sqrt{\sum_{1\le i,j\le p}(\widehat{\boldsymbol{\Sigma}}_{ij} - \overline{\boldsymbol{\Sigma}}_{ij})^2}$. We report the mean and standard deviation (in parenthesis) from 100 replications. The pair $(a, b) = (0, 0)$ corresponds to the i.i.d. case and is presented as the benchmark. The remaining (a, b) pairs correspond to nontrivial BEKK cases

		i.i.d.	BEKK		
	(a, b)	$(0, 0)$	$(0.15, 0.25)$	$(0.1, 0.65)$	$(0.05, 0.9)$
$(p, n) = (100, 125)$	original NLS	5.079	6.668	7.933	12.810
		(0.048)	(0.558)	(0.894)	(1.824)
	TV-adj NLS	5.080	5.219	5.308	6.433
		(0.049)	(0.096)	(0.078)	(0.711)
$(p, n) = (500, 625)$	original NLS	11.314	14.850	17.878	31.384
		(0.021)	(0.664)	(0.866)	(2.405)
	TV-adj NLS	11.320	11.362	11.469	11.970
		(0.026)	(0.028)	(0.050)	(0.196)

[3]The function “nlshrink_cov” in R package “nlshrink” is used in computing the nonlinear shrinkage estimator of the covariance matrix.

$(a, b) = (0.05, 0.9)$, because $a + b$ is close to one, the error of TV-adj NLS is larger. However, when p grows, the performance becomes closer to that of the NLS under the i.i.d. case.

3.2.2. *Performance under more choices of* (a, b). In this subsection, we present the results for a grid of (a, b)'s in the region $\{(a, b) : 0.05 \le a \le 0.5, 0.05 \le b \le 0.90, a + b \le 0.95\}$.

Empirical spectral distribution of the sample covariance matrices. In Figure 4, we plot the average Euclidean distance between the eigenvalues of $\mathbf{S}_n^0$ and $\widetilde{\mathbf{S}}_n$ or $\mathbf{S}_n$. We see that the Euclidean distance between the ESD of the original sample covariance matrix and that under the i.i.d. case grows substantially as a and $a + b$ increase. By contrast, the distance of the eigenvalues of the TV-adj sample covariance matrix to that under the i.i.d. case is close to zero for various (a, b) settings. The distance surface for the TV-adj sample covariance matrix is almost flat, except when $a + b$ approaches one, but when the dimension p increases, it again becomes flatter and closer to zero.

Population spectrum estimation. In Figure 5, we plot the average Euclidean distance between the estimated eigenvalues and the true eigenvalues. We compare the original NLS estimator and the proposed TV-adj NLS estimator. We see that the error of the original NLS estimator increases sharply as a and $a + b$ increase. It also gets larger when the dimension is higher. By contrast, the TV-adj NLS performs robustly well for various (a, b) settings and it dominantly outperforms the original NLS in all cases. The error surface for the TV-adj NLS is almost flat except when $a + b$ approaches one, but it gets closer to zero when p grows.

Unconditional covariance matrix estimation. Finally, in Figure 6, we plot the average Frobenius error of the NLS and TV-adj NLS in estimating the unconditional covariance matrix. We see that the original NLS estimator performs poorly when (a, b) deviates from $(0, 0)$. When p grows, the error also becomes larger. The TV-adj NLS dominantly outperforms the NLS with a lower estimation error in all cases. The error surface for the TV-adj NLS is almost flat and only slightly higher near the edge when $a + b$ is close to one.

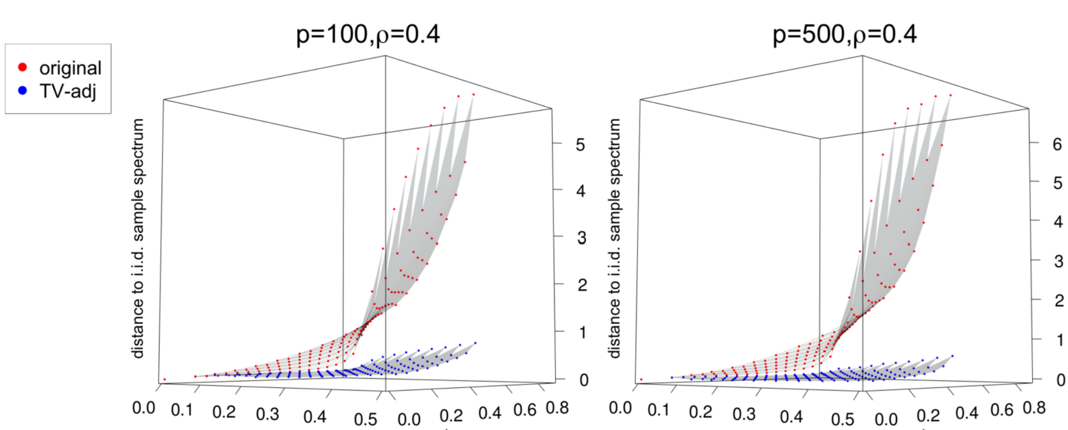

FIG. 4. *Euclidean distance between the eigenvalues of the sample covariance matrix/TV-adj sample covariance matrix under the BEKK model and the eigenvalues of the sample covariance matrix under the i.i.d. case for* $p = 100$ (*left*) *and* $p = 500$ (*right*). *The unconditional covariance matrix is* $\overline{\mathbf{\Sigma}} = (0.4)^{|i-j|}$. *The evaluation is made for a grid of* (a, b)*'s in the region* $\{(a, b) : 0.05 \le a \le 0.5, 0.05 \le b \le 0.90, a + b \le 0.95\}$.

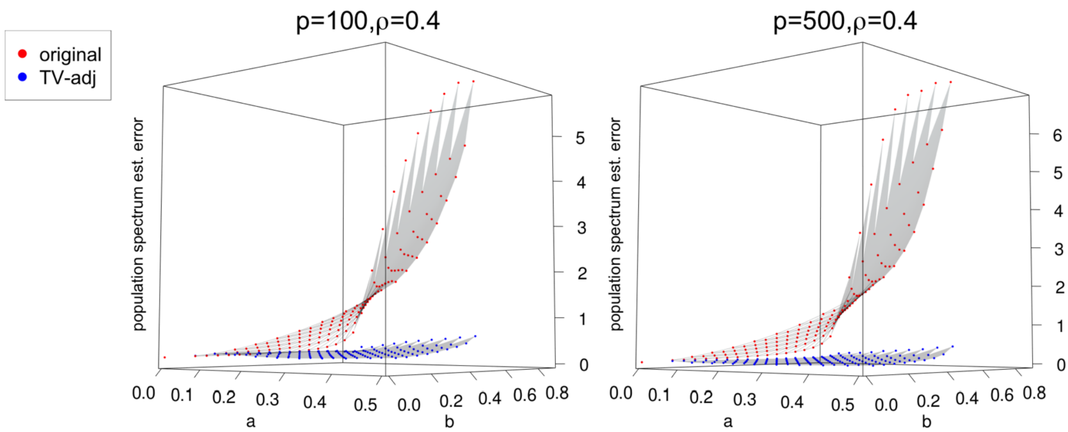

FIG. 5. *Estimation error of the population eigenvalues for* $p = 100$ *(left) and* $p = 500$ *(right). The unconditional covariance matrix is* $\overline{\mathbf{\Sigma}} = (0.4)^{|i-j|}$. *We compare the original NLS-spectrum estimator and the TV-adj NLS-spectrum estimator. The evaluation is made for a grid of* (a, b)*'s in the region* $\{(a, b) : 0.05 \leq a \leq 0.5, 0.05 \leq b \leq 0.90, a + b \leq 0.95\}$.

4. Conclusion. We investigate the limiting spectral properties of high-dimensional sample covariance matrices under large BEKK models. We show that dynamic (co)volatilities can impact the asymptotics and find the explicit condition under which the LSD is different from the i.i.d. case. To eliminate the impact, we propose a way to reverse transform the observations and show that the resulting sample covariance matrix has the same LSD as the i.i.d. case. Based on such a result, we develop consistent estimators of the spectral distribution and the oracle nonlinear shrinkage estimator of the unconditional covariance matrix.

5. Proof of Theorem 3. We divide the proof of Theorem 3 into three steps. In the first step, we show that replacing $(\widehat{a}_p, \widehat{b}_p)$ with (a_p, b_p) in equation (2.6) does not change the LSD of the TV-adj sample covariance matrices. In the second step, we show that the reverse transformation defined in (2.7) asymptotically restores i.i.d.ness up to a suitable orthogonal transformation. In the last step, we find the right orthogonal transformation.

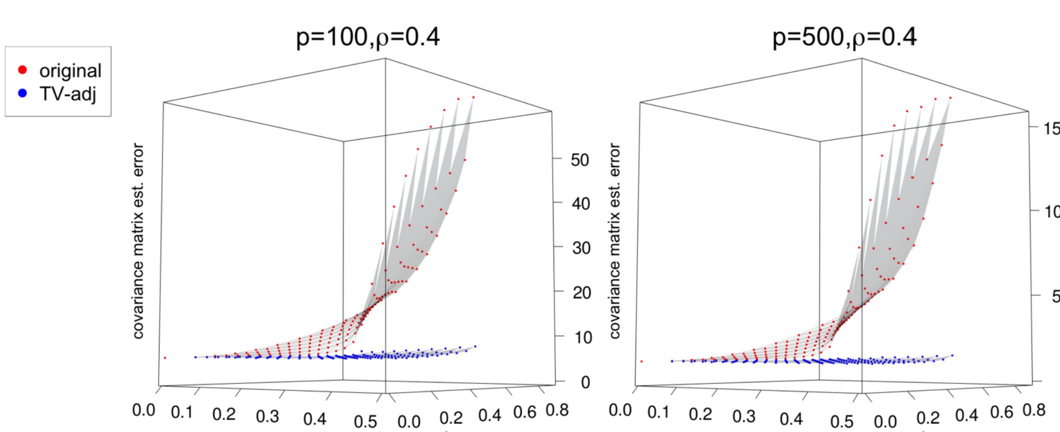

FIG. 6. *Estimation error of the unconditional covariance matrix in Frobenius norm. The unconditional covariance matrix is* $\overline{\mathbf{\Sigma}} = (0.4)^{|i-j|}$. *We compare the original NLS estimator and the TV-adj NLS estimator. The evaluation is made for a grid of* (a, b)*'s in the region* $\{(a, b) : 0.05 \leq a \leq 0.5, 0.05 \leq b \leq 0.90, a + b \leq 0.95\}$.

Step one: Define

$$\check{\mathbf{P}}_t = \frac{1-a_p-b_p+a_p b_p^{K_p}}{1-b_p}\mathbf{I} + \sum_{j=1}^{K_p} a_p b_p^{j-1}\mathbf{R}_{t-j}\mathbf{R}_{t-j}^T,$$

$$\check{\mathbf{R}}_t = (\check{\mathbf{P}}_t)^{-1/2}\mathbf{R}_t \quad \text{and}$$

$$\check{\mathbf{S}}_n = \frac{1}{n}\sum_{t=1}^{n}\check{\mathbf{R}}_t\check{\mathbf{R}}_t^T.$$

By Corollary A.42 of Bai and Silverstein (2010),

$$L^4\big(F^{\widetilde{\mathbf{S}}_n}, F^{\check{\mathbf{S}}_n}\big) \le \frac{2}{p}\operatorname{tr}(\widetilde{\mathbf{S}}_n + \check{\mathbf{S}}_n)\cdot\frac{1}{pn}\operatorname{tr}\big((\widetilde{\mathbf{R}}-\check{\mathbf{R}})(\widetilde{\mathbf{R}}-\check{\mathbf{R}})^T\big), \tag{5.1}$$

where $\widetilde{\mathbf{R}} = (\widetilde{\mathbf{R}}_1, \dots, \widetilde{\mathbf{R}}_n)$, and $\check{\mathbf{R}} = (\check{\mathbf{R}}_1, \dots, \check{\mathbf{R}}_n)$. Under condition (i) of Theorem 3, by Theorems 2.1 and 2.2 of Francq and Zakoïan (2004) and Theorem 2 and Corollary 3 of Francq and Zakoian (2007),

$$\widehat{a}_p - a_p = O_p\left(\frac{1}{\sqrt{n}}\right) \quad \text{and} \quad \widehat{b}_p - b_p = O_p\left(\frac{1}{\sqrt{n}}\right). \tag{5.2}$$

Under condition (ii) in Theorem 3, by Proposition 1, there exists a local MLE that satisfies (2.5). Note that under both conditions, $1-b_p \asymp 1-a_p-b_p$. Therefore, with probability tending one, for all n large enough, $((1-\hat{a}_p-\hat{b}_p)/(1-\hat{b}_p)) > \kappa > 0$. We have

$$P\left(\mathbf{P}_t = \frac{1-\widehat{a}_p-\widehat{b}_p+\widehat{a}_p\widehat{b}_p^{K_p}}{1-\widehat{b}_p}\mathbf{I} + \sum_{j=1}^{K_p}\widehat{a}_p\widehat{b}_p^{j-1}\mathbf{R}_{t-j}\mathbf{R}_{t-j}^T, \text{ for all } t\ge 1\right) \to 1. \tag{5.3}$$

Write

$$\varepsilon_n = \max\left(\max_{1\le j\le K_p}\left(\left|\frac{a_p b_p^{j-1}}{\widehat{a}_p\widehat{b}_p^{j-1}}-1\right|, \left|\frac{\widehat{a}_p\widehat{b}_p^{j-1}}{a_p b_p^{j-1}}-1\right|\right),\right.$$
$$\left.\left|\frac{(1-\widehat{a}_p-\widehat{b}_p+\widehat{a}_p\widehat{b}_p^{K_p})(1-b_p)}{(1-a_p-b_p+a_p b_p^{K_p})(1-\widehat{b}_p)}-1\right|, \left|\frac{(1-a_p-b_p+a_p b_p^{K_p})(1-\widehat{b}_p)}{(1-\widehat{a}_p-\widehat{b}_p+\widehat{a}_p\widehat{b}_p^{K_p})(1-b_p)}-1\right|\right).$$

By the assumption that $K_p \to \infty$ under condition (i), and $p^{\upsilon}(1-a_p-b_p) \to \infty$, $K_p \gg p^{\upsilon}\log p$ under condition (ii), we have

$$b_p^{K_p} = o(1-a_p-b_p). \tag{5.4}$$

Using (5.2) and the assumptions that $b_p p^{\varpi} \to \infty$, $K_p \ll p^{1/2-\varpi}$ under condition (i), or (2.5) and $K_p \ll p^{1/2-\varepsilon}$ under condition (ii), we get

$$\varepsilon_n = O_p\left((K_p+1)\left|\frac{\widehat{b}_p}{b_p}-1\right| + \left|\frac{\widehat{a}_p}{a_p}-1\right| + \frac{|\widehat{a}_p-a_p|+|\widehat{b}_p-b_p|}{1-a_p-b_p}\right)$$
$$= \begin{cases} O_p\left(\dfrac{K_p+1}{b_p p^{1/2}} + \dfrac{1}{a_p p^{1/2}}\right) = o_p(1) & \text{under condition (i)},\\[2ex] O_p\left(\dfrac{K_p+1}{p^{1/2-\varepsilon}} + \dfrac{1}{\min(a_p, 1-a_p-b_p)p^{1/2-\varepsilon}}\right) = o_p(1) & \text{under condition (ii)}.\end{cases} \tag{5.5}$$

Note that when $\varepsilon_n < 1$, under the event defined in (5.3), we have

$$0 < (1-\varepsilon_n)\check{\mathbf{P}}_t \le \mathbf{P}_t \le (1+\varepsilon_n)\check{\mathbf{P}}_t,$$

hence

$$\frac{1}{1+\varepsilon_n}\check{\mathbf{P}}_t^{-1} \le \mathbf{P}_t^{-1} \le \frac{1}{1-\varepsilon_n}\check{\mathbf{P}}_t^{-1}. \tag{5.6}$$

Because $(1-b_p) \asymp (1-a_p-b_p)$, we have $\check{\mathbf{P}}_t \ge (1-a_p-b_p)/(1-b_p)\mathbf{I} \ge c\mathbf{I}$ for some constant $c > 0$. By the definition of $\mathbf{P}_t$ in (2.6), $\mathbf{P}_t > \kappa\mathbf{I}$. Hence

$$\|\check{\mathbf{P}}_t^{-1/2}\| \le \big(\|\check{\mathbf{P}}_t^{-1}\|\big)^{1/2} \le \frac{1}{\sqrt{c}}, \tag{5.7}$$

and

$$\|\mathbf{P}_t^{-1/2}\| \le \big(\|\mathbf{P}_t^{-1}\|\big)^{1/2} \le \frac{1}{\sqrt{\kappa}}. \tag{5.8}$$

Using the fact that if $\mathbf{A} \ge 0$ and $\mathbf{B} \ge 0$, then $\operatorname{tr}(\mathbf{AB}) \ge 0$, we have, for all t,

$$\operatorname{tr}(\check{\boldsymbol{\Sigma}}_t) \le \frac{1}{c}\operatorname{tr}(\boldsymbol{\Sigma}_t), \qquad \operatorname{tr}(\widetilde{\boldsymbol{\Sigma}}_t) \le \frac{1}{\kappa}\operatorname{tr}(\boldsymbol{\Sigma}_t),$$

and

$$\operatorname{tr}\big(\check{\mathbf{R}}_t\check{\mathbf{R}}_t^T\big) \le \frac{1}{c}\operatorname{tr}\big(\mathbf{R}_t\mathbf{R}_t^T\big), \qquad \operatorname{tr}\big(\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t^T\big) \le \frac{1}{\kappa}\operatorname{tr}\big(\mathbf{R}_t\mathbf{R}_t^T\big).$$

Therefore,

$$\operatorname{tr}(\check{\mathbf{S}}_n) \le \frac{1}{c}\operatorname{tr}(\mathbf{S}_n) \quad \text{and} \quad \operatorname{tr}(\widetilde{\mathbf{S}}_n) \le \frac{1}{\kappa}\operatorname{tr}(\mathbf{S}_n).$$

By the independence between $(\mathbf{z}_t)$ and $(\boldsymbol{\Sigma}_t)$ and Assumption 1(iii), we have $E(\operatorname{tr}(\mathbf{S}_n)) = E(\mathbf{R}_t^T\mathbf{R}_t) = E(\mathbf{z}_t^T\boldsymbol{\Sigma}_t\mathbf{z}_t) = E(\operatorname{tr}(\boldsymbol{\Sigma}_t)) = \operatorname{tr}(\overline{\boldsymbol{\Sigma}}) = O(p)$. It follows that

$$\operatorname{tr}(\check{\mathbf{S}}_n) = O_p(p) \quad \text{and} \quad \operatorname{tr}(\widetilde{\mathbf{S}}_n) = O_p(p). \tag{5.9}$$

By (5.6) and the Löwner–Heinz inequality,

$$\frac{1}{\sqrt{1+\varepsilon_n}}\check{\mathbf{P}}_t^{-1/2} \le \mathbf{P}_t^{-1/2} \le \frac{1}{\sqrt{1-\varepsilon_n}}\check{\mathbf{P}}_t^{-1/2}.$$

By Weyl's theorem, we get that

$$\left\|\mathbf{P}_t^{-1/2} - \frac{1}{\sqrt{1+\varepsilon_n}}\check{\mathbf{P}}_t^{-1/2}\right\| \le \left(\frac{1}{\sqrt{1-\varepsilon_n}} - \frac{1}{\sqrt{1+\varepsilon_n}}\right) \cdot \|\check{\mathbf{P}}_t^{-1/2}\|. \tag{5.10}$$

By the triangle inequality, (5.5), (5.7) and (5.10), we get that

$$\begin{aligned}
\|\mathbf{P}_t^{-1/2} - \check{\mathbf{P}}_t^{-1/2}\| &\le \left\|\mathbf{P}_t^{-1/2} - \frac{1}{\sqrt{1+\varepsilon_n}}\check{\mathbf{P}}_t^{-1/2}\right\| + \left(1 - \frac{1}{\sqrt{1+\varepsilon_n}}\right) \cdot \|\check{\mathbf{P}}_t^{-1/2}\| \\
&= O_p\left(\frac{1}{\sqrt{1-\varepsilon_n}} - \frac{1}{\sqrt{1+\varepsilon_n}}\right) + O_p\left(1 - \frac{1}{\sqrt{1+\varepsilon_n}}\right) \\
&= O_p(\varepsilon_n) = o_p(1).
\end{aligned} \tag{5.11}$$

Moreover,

$$\begin{aligned}
&\operatorname{tr}\big((\widetilde{\mathbf{R}}_t - \check{\mathbf{R}}_t)(\widetilde{\mathbf{R}}_t - \check{\mathbf{R}}_t)^T\big) \\
&\quad = \mathbf{R}_t^T\big(\mathbf{P}_t^{-1/2} - \check{\mathbf{P}}_t^{-1/2}\big)^2\mathbf{R}_t \\
&\quad \le \|\mathbf{P}_t^{-1/2} - \check{\mathbf{P}}_t^{-1/2}\|^2 \cdot \|\mathbf{R}_t\|^2 = \|\mathbf{P}_t^{-1/2} - \check{\mathbf{P}}_t^{-1/2}\|^2 \cdot \big(\mathbf{z}_t^T\boldsymbol{\Sigma}_t\mathbf{z}_t\big).
\end{aligned}$$

By (5.3) and (5.11), we then get that

$$\operatorname{tr}\big((\widetilde{\mathbf{R}}_t - \check{\mathbf{R}}_t)(\widetilde{\mathbf{R}}_t - \check{\mathbf{R}}_t)^T\big) = o_p(p). \tag{5.12}$$

By (5.7), (5.8), (5.12) and the dominated convergence theorem,

$$\frac{1}{p} E\big(\operatorname{tr}\big((\widetilde{\mathbf{R}}_t - \check{\mathbf{R}}_t)(\widetilde{\mathbf{R}}_t - \check{\mathbf{R}}_t)^T\big)\big) = o(1),$$

hence

$$\frac{1}{pn} E\big(\operatorname{tr}\big((\widetilde{\mathbf{R}} - \check{\mathbf{R}})(\widetilde{\mathbf{R}} - \check{\mathbf{R}})^T\big)\big) = \frac{1}{p} E\big(\operatorname{tr}\big((\widetilde{\mathbf{R}}_t - \check{\mathbf{R}}_t)(\widetilde{\mathbf{R}}_t - \check{\mathbf{R}}_t)^T\big)\big) = o(1).$$

By Markov's inequality, we get

$$\frac{1}{pn} \operatorname{tr}\big((\widetilde{\mathbf{R}} - \check{\mathbf{R}})(\widetilde{\mathbf{R}} - \check{\mathbf{R}})^T\big) = o_p(1). \tag{5.13}$$

By (5.1), (5.9) and (5.13), we have

$$L\big(F^{\widetilde{\mathbf{S}}_n}, F^{\check{\mathbf{S}}_n}\big) = o_p(1). \tag{5.14}$$

Step two: We denote by $\mathcal{F}_t$ the σ-algebra generated by $\{\mathbf{z}_s, \infty < s \leq t\}$. For a $p \times p$ matrix $\mathcal{O}_t$ to be determined, which satisfies that

$$\mathcal{O}_t \text{ is } \mathcal{F}_{t-1}\text{-measurable} \quad \text{and} \quad \mathcal{O}_t\mathcal{O}_t^T = \mathcal{O}_t^T\mathcal{O}_t = \mathbf{I}, \tag{5.15}$$

we perform orthogonal transformation on $\mathbf{z}_t$ and get $\boldsymbol{\zeta}_t = \mathcal{O}_t\mathbf{z}_t$. We then define

$$\boldsymbol{\mathcal{R}}_t^0 = \overline{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{\zeta}_t, \qquad \boldsymbol{\mathcal{R}}^0 = (\boldsymbol{\mathcal{R}}_1^0, \dots, \boldsymbol{\mathcal{R}}_n^0) \quad \text{and} \quad \boldsymbol{\mathcal{S}}_n^0 = \frac{1}{n}\sum_{t=1}^n \boldsymbol{\mathcal{R}}_t^0(\boldsymbol{\mathcal{R}}_t^0)^T. \tag{5.16}$$

By (5.15) and the assumption that $\mathbf{z}_t \underset{\text{i.i.d.}}{\sim} N(0, \mathbf{I})$, we have

$$\boldsymbol{\zeta}_t \underset{\text{i.i.d.}}{\sim} N(0, \mathbf{I}). \tag{5.17}$$

By Theorem 1 of Marčenko and Pastur (1967), $F^{\mathbf{S}_n^0} \xrightarrow{P} F$, and $F^{\boldsymbol{\mathcal{S}}_n^0} \xrightarrow{P} F$. Hence,

$$L\big(F^{\mathbf{S}_n^0}, F^{\boldsymbol{\mathcal{S}}_n^0}\big) = o_p(1). \tag{5.18}$$

By (5.14), (5.18) and the triangle inequality, to show Theorem 3, it suffices to show that

$$L\big(F^{\boldsymbol{\mathcal{S}}_n^0}, F^{\check{\mathbf{S}}_n}\big) = o_p(1). \tag{5.19}$$

By Corollary A.42 of Bai and Silverstein (2010) again,

$$L^4\big(F^{\check{\mathbf{S}}_n}, F^{\boldsymbol{\mathcal{S}}_n^0}\big) \leq \frac{2}{p}\operatorname{tr}\big(\check{\mathbf{S}}_n + \boldsymbol{\mathcal{S}}_n^0\big) \cdot \frac{1}{pn}\operatorname{tr}\big((\check{\mathbf{R}} - \boldsymbol{\mathcal{R}}^0)(\check{\mathbf{R}} - \boldsymbol{\mathcal{R}}^0)^T\big). \tag{5.20}$$

We have $E(\operatorname{tr}(\boldsymbol{\mathcal{S}}_n^0/p)) = \operatorname{tr}(\overline{\boldsymbol{\Sigma}})/p = O(1)$, hence

$$\frac{1}{p}\operatorname{tr}\big(\boldsymbol{\mathcal{S}}_n^0\big) = O_p(1). \tag{5.21}$$

Combining (5.9) and (5.21) yields

$$\frac{1}{p}\operatorname{tr}\big(\check{\mathbf{S}}_n + \boldsymbol{\mathcal{S}}_n^0\big) = O_p(1). \tag{5.22}$$

Define

$$\mathbf{Q}_t = \check{\mathbf{P}}_t^{-1/2} \boldsymbol{\Sigma}_t^{1/2}. \tag{5.23}$$

We have $\check{\mathbf{R}}_t = \mathbf{Q}_t \mathbf{z}_t$, and $\check{\boldsymbol{\Sigma}}_t = \mathbf{Q}_t \mathbf{Q}_t^T$. We will show that for some $\mathcal{O}_t$ satisfying (5.15),

$$\frac{1}{p} E\big(\operatorname{tr}\big((\mathbf{Q}_t \mathcal{O}_t^T - \overline{\boldsymbol{\Sigma}}^{1/2})(\mathbf{Q}_t \mathcal{O}_t^T - \overline{\boldsymbol{\Sigma}}^{1/2})^T\big)\big) = o(1). \tag{5.24}$$

Then by the facts that $\check{\boldsymbol{\Sigma}}_t$ and $\mathcal{O}_t$ are $\mathcal{F}_{t-1}$-measurable, we have

$$\begin{aligned}
&\frac{1}{np} E\big(\operatorname{tr}\big((\check{\mathbf{R}} - \mathcal{R}^0)(\check{\mathbf{R}} - \mathcal{R}^0)^T\big)\big) \\
&\quad = \frac{1}{p} E\big(\operatorname{tr}\big((\check{\mathbf{R}}_t - \mathcal{R}_t^0)(\check{\mathbf{R}}_t - \mathcal{R}_t^0)^T\big)\big) \\
&\quad = \frac{1}{p} E\big(\operatorname{tr}\big((\mathbf{Q}_t - \overline{\boldsymbol{\Sigma}}^{1/2} \mathcal{O}_t)\mathbf{z}_t \mathbf{z}_t^T (\mathbf{Q}_t - \overline{\boldsymbol{\Sigma}}^{1/2} \mathcal{O}_t)^T\big)\big) \\
&\quad = \frac{1}{p} E\big(\operatorname{tr}\big((\mathbf{Q}_t \mathcal{O}_t^T - \overline{\boldsymbol{\Sigma}}^{1/2})(\mathbf{Q}_t \mathcal{O}_t^T - \overline{\boldsymbol{\Sigma}}^{1/2})^T\big)\big) = o(1),
\end{aligned}$$

which implies that

$$\frac{1}{pn} \operatorname{tr}\big((\check{\mathbf{R}} - \mathcal{R}^0)(\check{\mathbf{R}} - \mathcal{R}^0)^T\big) = o_p(1). \tag{5.25}$$

The desired bound (5.19) then follows from (5.20), (5.22) and (5.25).

Step three: It remains to show that there exists $\mathcal{O}_t$ satisfying (5.15) and (5.24). Because $K_p \ll p$, with probability one, for all p large enough, $\operatorname{rank}(\sum_{j=1}^{K_p} b_p^{j-1} \mathbf{R}_{t-j} \mathbf{R}_{t-j}^T) = K_p$. Write

$$\sum_{j=1}^{K_p} b_p^{j-1} \mathbf{R}_{t-j} \mathbf{R}_{t-j}^T = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T,$$

where $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_{K_p})$ and $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_{K_p})$ are the nonzero eigenvalues and the corresponding eigenvectors of $\sum_{j=1}^{K_p} b_p^{j-1} \mathbf{R}_{t-j} \mathbf{R}_{t-j}^T$, respectively. Recall that

$$\begin{aligned}
\check{\mathbf{P}}_t &= \frac{1 - a_p - b_p + a_p b_p^{K_p}}{1 - b_p} \mathbf{I} + \sum_{j=1}^{K_p} a_p b_p^{j-1} \mathbf{R}_{t-j} \mathbf{R}_{t-j}^T \\
&= \frac{1 - a_p - b_p + a_p b_p^{K_p}}{1 - b_p} \mathbf{I} + a_p \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T.
\end{aligned}$$

We have

$$\begin{aligned}
\check{\mathbf{P}}_t^{-1/2} &= \mathbf{U} \bigg(a_p \boldsymbol{\Lambda} + \frac{1 - a_p - b_p + a_p b_p^{K_p}}{1 - b_p} \mathbf{I} \bigg)^{-1/2} \mathbf{U}^T \\
&\quad + \sqrt{\frac{1 - b_p}{1 - a_p - b_p + a_p b_p^{K_p}}} (\mathbf{I} - \mathbf{U}\mathbf{U}^T).
\end{aligned} \tag{5.26}$$

By (1.1), we have

$$\boldsymbol{\Sigma}_t = \frac{1 - a_p - b_p}{1 - b_p} \overline{\boldsymbol{\Sigma}} + \sum_{s=1}^{\infty} a_p b_p^{s-1} \mathbf{R}_{t-s} \mathbf{R}_{t-s}^T, \tag{5.27}$$

which can be rewritten as

$$\boldsymbol{\Sigma}_t = \left(\frac{1-a_p-b_p+a_pb_p^{K_p}}{1-b_p}\overline{\boldsymbol{\Sigma}} + a_p\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T\right) + \left(\sum_{j=K_p+1}^{\infty} a_pb_p^{j-1}\mathbf{R}_{t-j}\mathbf{R}_{t-j}^T - \frac{a_pb_p^{K_p}}{1-b_p}\overline{\boldsymbol{\Sigma}}\right)$$
$$=: I_t + II_t.$$

Define

$$\widetilde{\boldsymbol{\Sigma}}_t = \mathbf{P}_t^{-1/2}\boldsymbol{\Sigma}_t\mathbf{P}_t^{-1/2} \quad \text{and} \quad \check{\boldsymbol{\Sigma}}_t = \check{\mathbf{P}}_t^{-1/2}\boldsymbol{\Sigma}_t\check{\mathbf{P}}_t^{-1/2}.$$

We have

$$\check{\boldsymbol{\Sigma}}_t = \check{\mathbf{P}}_t^{-1/2}I_t\check{\mathbf{P}}_t^{-1/2} + \check{\mathbf{P}}_t^{-1/2}II_t\check{\mathbf{P}}_t^{-1/2}.$$

By (5.26),

$$\check{\mathbf{P}}_t^{-1/2}I_t\check{\mathbf{P}}_t^{-1/2} = \overline{\boldsymbol{\Sigma}} - \overline{\boldsymbol{\Sigma}}\mathbf{U}\mathbf{U}^T - \mathbf{U}\mathbf{U}^T\overline{\boldsymbol{\Sigma}} + \mathbf{U}\mathbf{U}^T\overline{\boldsymbol{\Sigma}}\mathbf{U}\mathbf{U}^T + \mathcal{E}_t, \tag{5.28}$$

where

$$\mathcal{E}_t = \frac{1-a_p-b_p+a_pb_p^{K_p}}{1-b_p}\mathbf{U}\left(a_p\boldsymbol{\Lambda} + \frac{1-a_p-b_p+a_pb_p^{K_p}}{1-b_p}\mathbf{I}\right)^{-1/2}$$
$$\cdot \mathbf{U}^T\overline{\boldsymbol{\Sigma}}\mathbf{U}\left(a_p\boldsymbol{\Lambda} + \frac{1-a_p-b_p+a_pb_p^{K_p}}{1-b_p}\mathbf{I}\right)^{-1/2}\mathbf{U}^T$$
$$+ \sqrt{\frac{1-a_p-b_p+a_pb_p^{K_p}}{1-b_p}}(\mathbf{I}-\mathbf{U}\mathbf{U}^T)\overline{\boldsymbol{\Sigma}}\mathbf{U}\left(a_p\boldsymbol{\Lambda} + \frac{1-a_p-b_p+a_pb_p^{K_p}}{1-b_p}\mathbf{I}\right)^{-1/2}\mathbf{U}^T$$
$$+ \sqrt{\frac{1-a_p-b_p+a_pb_p^{K_p}}{1-b_p}}\mathbf{U}\left(a_p\boldsymbol{\Lambda} + \frac{1-a_p-b_p+a_pb_p^{K_p}}{1-b_p}\mathbf{I}\right)^{-1/2}\mathbf{U}^T\overline{\boldsymbol{\Sigma}}(\mathbf{I}-\mathbf{U}\mathbf{U}^T)$$
$$+ a_p\mathbf{U}\left(a_p\boldsymbol{\Lambda} + \frac{1-a_p-b_p+a_pb_p^{K_p}}{1-b_p}\mathbf{I}\right)^{-1/2}$$
$$\cdot \boldsymbol{\Lambda}\left(a_p\boldsymbol{\Lambda} + \frac{1-a_p-b_p+a_pb_p^{K_p}}{1-b_p}\mathbf{I}\right)^{-1/2}\mathbf{U}^T$$
$$=: \mathcal{E}_{1t} + \mathcal{E}_{2t} + \mathcal{E}_{3t} + \mathcal{E}_{4t}.$$

Because $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, we have

$$\operatorname{tr}(\mathcal{E}_{2t}) = \operatorname{tr}(\mathcal{E}_{3t}) = 0.$$

Moreover, because $K_p \ll p$ and $\|\overline{\boldsymbol{\Sigma}}\| = O(1)$, we have

$$0 \le \operatorname{tr}(\mathbf{U}^T\overline{\boldsymbol{\Sigma}}\mathbf{U}) = \sum_{i=1}^{K_p}\mathbf{u}_i^T\overline{\boldsymbol{\Sigma}}\mathbf{u}_i \le K_p\|\overline{\boldsymbol{\Sigma}}\| = o(p). \tag{5.29}$$

Furthermore, by $K_p \ll p$ and the fact that if $\mathbf{A} \ge 0$ and $\mathbf{B} \ge 0$, then $\operatorname{tr}(\mathbf{AB}) \ge 0$, we have

$$0 \le \operatorname{tr}(\mathcal{E}_{1t}) \le \operatorname{tr}(\mathbf{U}^T\overline{\boldsymbol{\Sigma}}\mathbf{U}) = o(p),$$
$$0 \le \operatorname{tr}(\mathcal{E}_{4t}) = \operatorname{tr}\left(a_p\boldsymbol{\Lambda}\left(a_p\boldsymbol{\Lambda} + \frac{1-a_p-b_p+a_pb_p^{K_p}}{1-b_p}\mathbf{I}\right)^{-1}\right) \le K_p = o(p).$$

Combining the results above yields

$$\operatorname{tr}(\mathcal{E}_t) = o(p).$$

Moreover, by (5.29) and that $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, we have

$$0 \le \operatorname{tr}\big(\mathbf{U}\mathbf{U}^T\overline{\boldsymbol{\Sigma}}\mathbf{U}\mathbf{U}^T\big) = \operatorname{tr}\big(\mathbf{U}\mathbf{U}^T\overline{\boldsymbol{\Sigma}}\big) = \operatorname{tr}\big(\overline{\boldsymbol{\Sigma}}\mathbf{U}\mathbf{U}^T\big) = \operatorname{tr}\big(\mathbf{U}^T\overline{\boldsymbol{\Sigma}}\mathbf{U}\big) = o(p).$$

Plugging the bounds above into (5.28) yields

$$\operatorname{tr}\big(\check{\mathbf{P}}_t^{-1/2} I_t \check{\mathbf{P}}_t^{-1/2}\big) - \operatorname{tr}(\overline{\boldsymbol{\Sigma}}) = o(p). \tag{5.30}$$

About term $\check{\mathbf{P}}_t^{-1/2} II_t \check{\mathbf{P}}_t^{-1/2}$, because $\check{\mathbf{P}}_t \ge ((1-a_p-b_p)/(1-b_p))\mathbf{I}$, we have

$$\begin{aligned}
&\big|\operatorname{tr}\big(\check{\mathbf{P}}_t^{-1/2} II_t \check{\mathbf{P}}_t^{-1/2}\big)\big| \\
&\quad\le \frac{a_p b_p^{K_p}}{1-a_p-b_p}\operatorname{tr}(\overline{\boldsymbol{\Sigma}}) + \frac{b_p^{K_p}(1-b_p)}{1-a_p-b_p}\operatorname{tr}\Bigg(\sum_{s=1}^{\infty} a_p b_p^{s-1}\mathbf{R}_{t-K_p-s}\mathbf{R}_{t-K_p-s}^T\Bigg) \\
&\quad\le \frac{a_p b^{K_p}}{1-a_p-b_p}\operatorname{tr}(\overline{\boldsymbol{\Sigma}}) + \frac{b_p^{K_p}(1-b_p)}{1-a_p-b_p}\operatorname{tr}(\boldsymbol{\Sigma}_{t-K_p}).
\end{aligned}$$

We have $E(\operatorname{tr}(\boldsymbol{\Sigma}_{t-K_p})) = \operatorname{tr}(\overline{\boldsymbol{\Sigma}}) = O(p)$, hence $\operatorname{tr}(\boldsymbol{\Sigma}_{t-K_p}) = O_p(p)$. By (5.4), we have

$$\begin{aligned}
\big|\operatorname{tr}\big(\check{\mathbf{P}}_t^{-1/2} II_t \check{\mathbf{P}}_t^{-1/2}\big)\big| &= o(1)\big(\operatorname{tr}(\overline{\boldsymbol{\Sigma}}) + \operatorname{tr}(\boldsymbol{\Sigma}_{t-K_p})\big) \\
&= o_p(p).
\end{aligned} \tag{5.31}$$

By (5.30) and (5.31),

$$\frac{1}{p}\big|\operatorname{tr}(\check{\boldsymbol{\Sigma}}_t) - \operatorname{tr}(\overline{\boldsymbol{\Sigma}})\big| = o_p(1). \tag{5.32}$$

We now define $\mathcal{O}_t$ that satisfies (5.15) and (5.24). Let $\mathbf{G}_t = \sqrt{(1-a_p-b_p)/(1-b_p)} \times \check{\mathbf{P}}_t^{-1/2}\overline{\boldsymbol{\Sigma}}^{1/2}$. We have $\mathbf{G}_t\mathbf{G}_t^T = ((1-a_p-b_p)/(1-b_p))\check{\mathbf{P}}_t^{-1/2}\overline{\boldsymbol{\Sigma}}\check{\mathbf{P}}_t^{-1/2}$. By (5.27), $\boldsymbol{\Sigma}_t \ge ((1-a_p-b_p)/(1-b_p))\overline{\boldsymbol{\Sigma}}$. Hence

$$\check{\boldsymbol{\Sigma}}_t \ge \mathbf{G}_t\mathbf{G}_t^T. \tag{5.33}$$

Define

$$\mathcal{Q}_t = \mathbf{G}_t\big(\mathbf{I} + \mathbf{G}_t^{-1}(\check{\boldsymbol{\Sigma}}_t - \mathbf{G}_t\mathbf{G}_t^T)(\mathbf{G}_t^T)^{-1}\big)^{1/2},$$

and

$$\mathcal{O}_t = \mathcal{Q}_t^T(\mathbf{Q}_t^T)^{-1},$$

where, recall that, $\mathbf{Q}_t$ is defined in (5.23). By definition, $\mathcal{O}_t$ is $\mathcal{F}_{t-1}$-measurable. Moreover, it is straightforward to verify that $\mathcal{Q}_t\mathcal{Q}_t^T = \mathbf{Q}_t\mathbf{Q}_t^T = \check{\boldsymbol{\Sigma}}_t$, from which we get that $\mathcal{O}_t\mathcal{O}_t^T = \mathcal{O}_t^T\mathcal{O}_t = \mathbf{I}$. Therefore, $\mathcal{O}_t$ satisfies (5.15).

It remains to show that $\mathcal{O}_t$ satisfies (5.24). By (5.26),

$$\check{\mathbf{P}}_t^{-1/2} \ge \sqrt{\frac{1-b_p}{1-a_p-b_p+a_p b_p^{K_p}}}(\mathbf{I} - \mathbf{U}\mathbf{U}^T). \tag{5.34}$$

By (5.33) and (5.34),

$$
\begin{aligned}
\operatorname{tr}\big(\mathcal{Q}_t^T\overline{\boldsymbol{\Sigma}}^{1/2}\big) &= \operatorname{tr}\big(\mathcal{Q}_t\overline{\boldsymbol{\Sigma}}^{1/2}\big) \\
&= \sqrt{\frac{1-a_p-b_p}{1-b_p}}\operatorname{tr}\big(\big(\mathbf{I}+\mathbf{G}_t^{-1}(\check{\boldsymbol{\Sigma}}_t-\mathbf{G}_t\mathbf{G}_t^T)(\mathbf{G}_t^T)^{-1}\big)^{1/2}\overline{\boldsymbol{\Sigma}}^{1/2}\check{\mathbf{P}}_t^{-1/2}\overline{\boldsymbol{\Sigma}}^{1/2}\big) \\
&\geq \sqrt{\frac{1-a_p-b_p}{1-b_p}}\operatorname{tr}\big(\check{\mathbf{P}}_t^{-1/2}\overline{\boldsymbol{\Sigma}}\big) \\
&\geq \sqrt{\frac{1-a_p-b_p}{1-a_p-b_p+a_pb_p^{K_p}}}\operatorname{tr}\big((\mathbf{I}-\mathbf{U}\mathbf{U}^T)\overline{\boldsymbol{\Sigma}}\big) \\
&= \operatorname{tr}(\overline{\boldsymbol{\Sigma}})+o(p),
\end{aligned}
\tag{5.35}
$$

where the last equation holds by (5.4) and (5.29). By the definition of $\mathcal{O}_t$, (5.32) and (5.35), we get that

$$
\begin{aligned}
0 &\leq \frac{1}{p}\operatorname{tr}\big((\mathbf{Q}_t\mathcal{O}_t^T-\overline{\boldsymbol{\Sigma}}^{1/2})(\mathbf{Q}_t\mathcal{O}_t^T-\overline{\boldsymbol{\Sigma}}^{1/2})^T\big) \\
&= \frac{1}{p}\operatorname{tr}\big((\mathbf{Q}_t-\overline{\boldsymbol{\Sigma}}^{1/2}\mathcal{O}_t)(\mathbf{Q}_t-\overline{\boldsymbol{\Sigma}}^{1/2}\mathcal{O}_t)^T\big) \\
&= \frac{1}{p}\operatorname{tr}\big((\mathcal{Q}_t-\overline{\boldsymbol{\Sigma}}^{1/2})(\mathcal{Q}_t-\overline{\boldsymbol{\Sigma}}^{1/2})^T\big) \\
&= \frac{1}{p}\big(\operatorname{tr}(\check{\boldsymbol{\Sigma}}_t)-\operatorname{tr}(\overline{\boldsymbol{\Sigma}})\big)+\frac{1}{p}\big(\operatorname{tr}(\overline{\boldsymbol{\Sigma}})-\operatorname{tr}\big(\mathcal{Q}_t\overline{\boldsymbol{\Sigma}}^{1/2}\big)\big)+\frac{1}{p}\big(\operatorname{tr}(\overline{\boldsymbol{\Sigma}})-\operatorname{tr}\big(\mathcal{Q}_t^T\overline{\boldsymbol{\Sigma}}^{1/2}\big)\big) \\
&\leq o_p(1).
\end{aligned}
$$

Hence, $\operatorname{tr}((\mathbf{Q}_t\mathcal{O}_t^T-\overline{\boldsymbol{\Sigma}}^{1/2})(\mathbf{Q}_t\mathcal{O}_t^T-\overline{\boldsymbol{\Sigma}}^{1/2})^T)/p=o_p(1)$. In addition, we have

$$
\begin{aligned}
&\frac{1}{p}\operatorname{tr}\big((\mathcal{Q}_t-\overline{\boldsymbol{\Sigma}}^{1/2})(\mathcal{Q}_t-\overline{\boldsymbol{\Sigma}}^{1/2})^T\big) \\
&\quad= \frac{1}{p}\sum_{1\leq i,j\leq p}\big((\mathcal{Q}_t)_{ij}-(\overline{\boldsymbol{\Sigma}}^{1/2})_{ij}\big)^2 \\
&\quad\leq \frac{2}{p}\sum_{1\leq i,j\leq p}\big((\mathcal{Q}_t)_{ij}\big)^2+\frac{2}{p}\sum_{1\leq i,j\leq p}\big((\overline{\boldsymbol{\Sigma}}^{1/2})_{ij}\big)^2 \\
&\quad= \frac{2}{p}\operatorname{tr}(\check{\boldsymbol{\Sigma}}_t)+\frac{2}{p}\operatorname{tr}(\overline{\boldsymbol{\Sigma}}).
\end{aligned}
$$

By (5.7) and Assumption 1(iii), we have $E(\operatorname{tr}(\check{\boldsymbol{\Sigma}}_t/p))\leq CE(\operatorname{tr}(\boldsymbol{\Sigma}_t/p))=O(\|\overline{\boldsymbol{\Sigma}}\|)=O(1)$. The bound (5.24) follows from the dominated convergence theorem.

Acknowledgments. The authors thank the editor and two anonymous referees for constructive comments and suggestions that led to significant improvement in the paper.

Funding. Research is supported in part by RGC grants GRF 16304019, GRF 15302321 and GRF 16304521 of the HKSAR, NSFC 72101226 of China, and SRG2023-00001-FBA and MYRG-GRG2023-00119-FBA of the University of Macau.

SUPPLEMENTARY MATERIAL

Supplement to "High-dimensional covariance matrices under dynamic volatility models: Asymptotics and shrinkage estimation" (DOI: 10.1214/24-AOS2381SUPP; .pdf). This supplement contains the proofs of Theorems 1, 2, 4 and 5, Corollary 1 and Proposition 1, and additional simulation results.

REFERENCES

AO, M., LI, Y. and ZHENG, X. (2019). Approaching mean-variance efficiency for large portfolios. *Rev. Financ. Stud.* **32** 2890–2919.

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

BANNA, M. and MERLEVÈDE, F. (2015). Limiting spectral distribution of large sample covariance matrices associated with a class of stationary processes. *J. Theoret. Probab.* **28** 745–783. MR3370674 https://doi.org/10.1007/s10959-013-0508-x

BHATTACHARJEE, M. and BOSE, A. (2016). Large sample behaviour of high dimensional autocovariance matrices. *Ann. Statist.* **44** 598–628. MR3476611 https://doi.org/10.1214/15-AOS1378

BOLLERSLEV, T., ENGLE, R. F. and WOOLDRIDGE, J. M. (1988). A capital asset pricing model with time-varying covariances. *J. Polit. Econ.* **96** 116–131.

DE NARD, G., LEDOIT, O. and WOLF, M. (2021). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *J. Financ. Econom.* **19** 236–257.

DING, Y., LI, Y. and ZHENG, X. (2021). High dimensional minimum variance portfolio estimation under statistical factor models. *J. Econometrics* **222** 502–515. MR4234830 https://doi.org/10.1016/j.jeconom.2020.07.013

DING, Y. and ZHENG, X. (2024). Supplement to "High-dimensional covariance matrices under dynamic volatility models: Asymptotics and shrinkage estimation." https://doi.org/10.1214/24-AOS2381SUPP

DING, Z. and ENGLE, R. F. (2001). Large scale conditional covariance matrix modeling, estimation and testing. *Academia Economic Papers* **29** 157–184.

EL KAROUI, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.* **36** 2757–2790. MR2485012 https://doi.org/10.1214/07-AOS581

ENGLE, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *J. Bus. Econom. Statist.* **20** 339–350. MR1939905 https://doi.org/10.1198/073500102288618487

ENGLE, R. F., GRANGER, C. W. J. and KRAFT, D. (1984). Combining competing forecasts of inflation using a bivariate ARCH model. *J. Econom. Dynam. Control* **8** 151–165. MR0781644 https://doi.org/10.1016/0165-1889(84)90031-9

ENGLE, R. F. and KRONER, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric Theory* **11** 122–150. MR1325104 https://doi.org/10.1017/S0266466600009063

ENGLE, R. F., LEDOIT, O. and WOLF, M. (2019). Large dynamic covariance matrices. *J. Bus. Econom. Statist.* **37** 363–375. MR3948411 https://doi.org/10.1080/07350015.2017.1345683

FRANCQ, C. and ZAKOÏAN, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA–GARCH processes. *Bernoulli* **10** 605–637. MR2076065 https://doi.org/10.3150/bj/1093265632

FRANCQ, C. and ZAKOIAN, J.-M. (2007). Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. *Stochastic Process. Appl.* **117** 1265–1284. MR2343939 https://doi.org/10.1016/j.spa.2007.01.001

JIN, B., WANG, C., MIAO, B. and LO HUANG, M.-N. (2009). Limiting spectral distribution of large-dimensional sample covariance matrices generated by VARMA. *J. Multivariate Anal.* **100** 2112–2125. MR2543090 https://doi.org/10.1016/j.jmva.2009.06.011

LEDOIT, O. and PÉCHÉ, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probab. Theory Related Fields* **151** 233–264. MR2834718 https://doi.org/10.1007/s00440-010-0298-3

LEDOIT, O. and WOLF, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Statist.* **40** 1024–1060. MR2985942 https://doi.org/10.1214/12-AOS989

LEDOIT, O. and WOLF, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *J. Multivariate Anal.* **139** 360–384. MR3349498 https://doi.org/10.1016/j.jmva.2015.04.006

LEDOIT, O. and WOLF, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Rev. Financ. Stud.* **30** 4349–4388.

LEDOIT, O. and WOLF, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *Ann. Statist.* **48** 3043–3065. MR4152634 https://doi.org/10.1214/19-AOS1921

LIU, H., AUE, A. and PAUL, D. (2015). On the Marčenko–Pastur law for linear time series. *Ann. Statist.* **43** 675–712. MR3319140 https://doi.org/10.1214/14-AOS1294

MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR, Sb.* **1** 457.

MERLEVÈDE, F. and PELIGRAD, M. (2016). On the empirical spectral distribution for matrices with long memory and independent rows. *Stochastic Process. Appl.* **126** 2734–2760. MR3522299 https://doi.org/10.1016/j.spa.2016.02.016

PAKEL, C., SHEPHARD, N., SHEPPARD, K. and ENGLE, R. F. (2021). Fitting vast dimensional time-varying covariance models. *J. Bus. Econom. Statist.* **39** 652–668. MR4272926 https://doi.org/10.1080/07350015.2020.1713795

PEDERSEN, R. S. and RAHBEK, A. (2014). Multivariate variance targeting in the BEKK–GARCH model. *Econom. J.* **17** 24–55. MR3171211 https://doi.org/10.1111/ectj.12019

SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. MR1370408 https://doi.org/10.1006/jmva.1995.1083

SILVERSTEIN, J. W. and BAI, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.* **54** 175–192. MR1345534 https://doi.org/10.1006/jmva.1995.1051

YANG, X., ZHENG, X. and CHEN, J. (2021). Testing high-dimensional covariance matrices under the elliptical distribution and beyond. *J. Econometrics* **221** 409–423. MR4215033 https://doi.org/10.1016/j.jeconom.2020.05.017

YAO, J. (2012). A note on a Marčenko–Pastur type theorem for time series. *Statist. Probab. Lett.* **82** 22–28. MR2863018 https://doi.org/10.1016/j.spl.2011.08.011

YASKOV, P. A. (2017). On the spectrum of sample covariance matrices for time series. *Teor. Veroyatn. Primen.* **62** 542–555. MR3684648 https://doi.org/10.1137/S0040585X97T988721

YIN, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.* **20** 50–68. MR0862241 https://doi.org/10.1016/0047-259X(86)90019-9

ZHENG, X. and LI, Y. (2011). On the estimation of integrated covariance matrices of high dimensional diffusion processes. *Ann. Statist.* **39** 3121–3151. MR3012403 https://doi.org/10.1214/11-AOS939