Forecasting High-Dimensional Covariance Matrices Using High-Dimensional Principal Component Analysis

Hideto Shigemoto [*], Takayuki Morimoto [†]

November 17, 2023

Abstract

We modify the recently proposed forecasting model of high-dimensional covariance matrices (HDCM) of asset returns using high-dimensional principal component analysis (PCA). It is well-known that when the sample size is smaller than the dimension, eigenvalues estimated by classical PCA have a bias. In particular, a very small number of eigenvalues are extremely large and they are called spiked eigenvalues. High-dimensional PCA gives eigenvalues which correct the biases of the spiked eigenvalues. This situation also happens in the financial field, especially in situations where high-frequency and high-dimensional data are handled. The research aims to estimate the HDCM of asset returns using high-dimensional PCA for the realized covariance matrix using the Nikkei 225 data, it estimates 5- and 10-minute intraday asset-returns intervals. We construct time-series models for eigenvalues which are estimated by each PCA, and forecast HDCM. Our simulation analysis shows that the high-dimensional PCA has better estimation performance than classical PCA for the estimating integrated covariance matrix. In our empirical analysis, we show that we will be able to improve the forecasting performance using the high-dimensional PCA and make a portfolio with smaller variance.

Keywords: covariance forecasting; high-dimensional covariance; principal component analysis; high-frequency data; time series

1 Introduction

Modeling and forecasting covariance matrices of asset returns have an essential role in portfolio allocations and risk management. For estimating and forecasting covariance matrix, a lot of papers are published on both low- and high-frequency data. Concerning the low-frequency data, the multivariate GARCH models [1], for example, BEKK-GARCH [2] and DCC-GARCH [3, 4], are usually used to estimate and forecast the covariance matrix as latent. On the other hand, the availability of high-frequency data recently enabled the direct estimation of the covariance

* Group Risk Management Department, Nomura Holdings, Inc., 2-2-2 Otemachi, Chiyoda-ku, Tokyo 100-8130, Japan.

† Corresponding author: Department of Mathematical Sciences, Kwansei Gakuin University, 1, Gakuen Uegahara, Sanda, Hyogo 669-1330, Japan.
morimot@kwansei.ac.jp

matrix, for example, the realized covariance matrix estimator [5], and the multivariate realized kernel estimator [6]. Additionally, some forecasting models such as the multivariate HAR [7], conditional autoregressive Wishart (CAW) [8], and realized DCC [9] models, use these covariance estimators to forecast them. However, when the dimensions increase, these covariance estimators and forecasting models have less accurate performance and suffer from an increase in the number of estimated parameters because of various reasons, such as the curse of dimensionality.

To solve these problems, the DCC-NL model which can overcome the curse of dimensionality using nonlinear shrinkage estimation is proposed [10]. To analyze the conditional high-dimensional covariance matrix (HDCM), recent studies using some multivariate GARCH models use the DCC-NL model instead of Tse and Tsui's and Engle's DCC-GARCH models [11, 12, 13, 14]. Then, to solve the curse of dimensionality, many studies assume that the covariance matrix process or the price process follows a factor structure. Wang and Zou [15] propose a covariance estimator assuming that the integrated covariance matrix is sparse. Considering a sparse covariance matrix allows only important elements to remain and also reduces the number of elements to be estimated. In addition, Tao et al. [16] introduce a covariance estimator which uses the matrix factor structure for an HDCM. We can obtain not only a consistent estimator of an HDCM but also a forecasted value using the vector autoregressive (VAR) model for a low-dimensional factor covariance matrix. Kim et al. [17] propose a threshold covariance estimator to regularize some realized covariance measures under the same assumption as [15]. Shen et al. [18] apply the method proposed by [16] to a realized covariance matrix and consider the CAW model instead of the VAR model for the factors. However, these studies assume sparsity in the integrated covariance matrix itself, which represents the target to be estimated. If there are some common factors across asset returns, the assumption that the integrated covariance is sparse becomes unrealistic because there are correlations among all pairs of assets through the common factors [19, 20, 21, 22].

Fan et al. [19] propose the principal orthogonal complement thresholding (POET) method which assumes sparsity, not for the covariance matrix itself, but for the covariance matrix of the residual process, and estimates the latent factor using principal component analysis (PCA) to solve some problems. For high-frequency data, Fan et al. [20] assume the observable factor structure inspired by [23], and propose the covariance estimator under the assumption that the covariance matrix of the residual process is sparse. To estimate the latent factor structure, Aït-Sahalia and Xiu [24] impose sparsity on the residual covariance matrix and apply POET to high-frequency data using PCA to estimate an HDCM. They show that even when the factor is latent, if the residual covariance matrix is sufficiently sparse, the factor part can be estimated by PCA on the consistent estimator of the integrated covariance matrix, like the realized covariance matrix. In addition, they show that their estimator is a consistent estimator even if the interval of intraday return is $\Delta \to 0$ and the dimension is $d \to \infty$. In addition, Dai et al. [25] also propose an estimation method of the sparse residual covariance matrix using thresholding, and a high-dimensional covariance estimator using the POET estimator. The difference between [24] and [25] is the sparse structure. While Aït-Sahalia and Xiu [24] assume the block-diagonalize structure instead of thresholding, Dai et al. [25] do not assume the block-diagonalize structure but set a more general assumption, and use soft-, hard-, and adaptive-lasso (AL) [26], and smoothly clipped absolute deviation (SCAD) [27] thresholding. For the sparse estimation of the residual covariance matrix, Cai and Liu [28] propose the adaptive and hard thresholding method, but this method cannot guarantee the positive definiteness under the finite sample [29], and also has less performance than [25]. Brownlees et al. [21] propose the realized network estimator using the graphical lasso to estimate the precision matrix. Jian et al. [29] build time-series models for estimated eigenvalues based on the estimator of [24], and forecast the HDCM. In addition, they propose the regularized method to guarantee the positive definiteness.

The classical PCA, which is used by these models, creates a bias under $d > M$; d is the dimension of a covariance matrix and M is the sample size [15, 30, 24, 31]. Wang and Fan [31] characterize the asymptotic distribution of empirical eigenvalues under the i.i.d setting and $d > M$. They also propose the shrinkage POET (SPOET) method based on their asymptotic distribution. The SPOET method corrects the biases of eigenvalues estimated by classical PCA.

In this paper, we estimate the HDCM under the factor structure for the high-frequency data, and create the forecasting models using its eigenvalues. It is well-known that the realized covariance matrix is a consistent estimator of the integrated covariance matrix when the number of intraday observations M goes to ∞. However, in the empirical situation, we consider the microstructure noise, and often use the realized covariance matrix which is estimated using 5- or 10-minute interval intraday returns. In this case, since the Japanese stock market opens from 9 a.m. to 3 p.m. with an hour break, the sample sizes are 60 and 30 per day. Under such a situation, although we want to consider a large portfolio including 100 or 200 stocks, the matrix dimension is larger than the sample size, $d > M$. Therefore, we apply spoet corresponding to $d > M$ to the realized covariance matrix, rather than the POET using PCA as considered in [24, 25]. Additionally, we construct the forecasting models similar to [29], by deriving the eigenvalues of the realized covariance matrix estimated using SPOET.

There are two contributions to the literature. First, this paper shows through a simulation study that SPOET considered in the i.i.d. setting has excellent performance for estimating the integrated covariance matrix under the assumption of continuous Itô semi-martingale. Second, our empirical analysis shows that the forecasting models using SPOET are more accurate covariance matrix than the models using the POET. Hence, using our proposed models gives us a more accurate covariance estimator under the high-dimensional setting that results in bad performance and unreliable results. This point is the largest difference between [29] and this paper. Although Jian et al. [29] do not consider the relationship between the dimension of the covariance matrix and the sample size of intraday, we focus on the relationship and make these models forecast more accurately than their models.

The paper is organized as follows: Section 2 explains the factor model, the sparse estimations, and the principal component analysis to estimate the factor part. Section 3 introduces the forecasting model of estimated eigenvalues by PCA used in the empirical analysis.

2 Factor Model and PCA

2.1 Factor Structure

We assume that the log-price Y follows a continuous-time factor model,

$$Y_t = \beta X_t + Z_t, \tag{1}$$

where Y_t is a d-dimensional vector process, X_t is a r-dimensional latent common factor process, Z_t is the d-dimensional idiosyncratic component, and β is a $d \times r$ constant-factor loading matrix. In addition, X_t and Z_t are independent. In this paper, the number of factors r is unknown. Here, we assume that X_t and Z_t are continuous Itô semi-martingale, as with [24, 25] as follows:

$$X_t = \int_0^t h_s ds + \int_0^t \eta_s dW_s, \;\; Z_t = \int_0^t f_s ds + \int_0^t \gamma_s dB_s.$$

Then, the integrated covariance matrices of X_t, Z_t, and Y_t are defined under Assumptions 1, 2, and 3, and the sparsity assumption of [25] as follows:

$$\Sigma_{X_t} = \int_0^t \eta_s \eta_s' ds, \quad \Sigma_{Z_t} = \int_0^t \gamma_s \gamma_s' ds,$$
$$\Sigma_{Y_t} = \beta \Sigma_{X_t} \beta' + \Sigma_{Z_t}. \tag{2}$$

Although Jian et al. [29] consider the factor model following Assumption 1, 2, 3, 4, and 5 of [24], we assume more general sparsity of [25] and we do not assume that idiosyncratic component is block diagonal.

2.2 Sparsity

To estimate an HDCM, a certain condition of sparsity is necessary for dimension reduction and factor model. However, the sparsity assumption of the covariance matrix itself is inappropriate from the viewpoint of the factor model. To solve this problem, we assume that the covariance matrix of the idiosyncratic component Σ_Z is sparse, and then the form of Equation (2) becomes a low-rank plus sparse structure. A low-rank plus sparsity structure of the residual covariance matrix turns out to be a good match for asset high-frequency data [24] and guarantees a well-conditioned estimator as well as its precision matrix [25].

We use four types of thresholding functions, hard-, soft-, adaptive lasso (AL) and smoothly clipped absolute deviation (SCAD) threshold, for Σ_Z as following:

$$s_\lambda^{\text{Hard}}(z) = z\mathbf{1}(|z| > \lambda), \;\; s_\lambda^{\text{Soft}}(z) = \text{sign}(z)(|z| - \lambda)_+, \;\; s_\lambda^{AL}(z) = \text{sign}(z)(|z| - \lambda^{\eta+1}|z|^{-\eta})_+,$$

$$s_\lambda^{SCAD}(z) = \begin{cases} \text{sign}(z)(|z| - \lambda)_+, & |z| \le 2\lambda; \\ \frac{(a-1)z - \text{sign}(z)a\lambda}{a-2}, & 2\lambda < |z| \le a\lambda; \\ z, & a\lambda < |z|. \end{cases}$$

where we set $a = 3.7$ and $\eta = 1$ same as [32]. We adopt these thresholding functions and estimate the residual covariance matrix as follows:

$$\tilde{\Sigma}^S_{Z_t,ij} = \begin{cases} \hat{\Sigma}_{Z_t,ij}, & i = j; \\ s_{\lambda_{ij}}(\hat{\Sigma}_{Z_t,ij}), & i \neq j. \end{cases}$$

Dai et al. [25] denote that despite these estimations lead to the same convergence rate from their analysis, the results of finite sample performance of the covariance matrix in their simulation study and empirical analysis are quite different.

2.2.1 Thresholding Method

Following [25], the thresholding λ_{ij} in sparse functions is estimated as follows:

$$\lambda_{ij} = \tau \sqrt{\hat{\Sigma}_{Z_t,ii} \hat{\Sigma}_{Z_t,jj}},$$

where τ is a constant to be determined. Under the finite sample, we use a grid search to guarantee positive semi-definite. We divide into K pieces in $\tau \in [0, 1]$ and gradually increase τ until the final high-dimensional covariance matrix becomes positive semi-definite. As τ becomes larger, the degree of sparsity of the residual covariance increases, and, finally, the matrix becomes a diagonal matrix [25]. Thus, an estimated HDCM always becomes positive semi-definite.

2.2.2 The Number of Factors

If the log-price is observed by latent common factors, we have to estimate the number of factors. The consistent estimator of the number of latent factors is proposed by [24] under the continuous-time setting without random matrix theory. We adopt their estimator, which minimizes the penalized function using an estimator of the integrated covariance matrix $\hat{\Sigma}_t$:

$$\hat{r}^t = \arg \min_{1 \leq j \leq r_{\max}} \left(\frac{\lambda_j(\hat{\Sigma}_{Y_t})}{d} + j \times g(M, d) \right) - 1, \tag{3}$$

where $r_{\max}$ is 20. In theory, the choice of $r_{\max}$ is not important. This is simply used to avoid making economically meaningless choice of r in finite samples [24]. The function $g(n, d)$ is defined as follows:

$$g(M, d) = 0.02 \times \hat{\lambda}^t_{\min\left(\frac{d}{2}, \frac{M}{2}\right)}(\hat{\Sigma}_{Y_t}) \left(\frac{\log d}{M} \right)^{\frac{1}{4}}. \tag{4}$$

2.3 PCA for High-Frequency Data

To estimate an HDCM, we show the PCA for the realized covariance matrix estimated by high-frequency data following [29]. Here, $y_{j,t}$ is the j-th intraday log-return observed on day t. The realized covariance matrix is defined as follows:

$$\hat{\Sigma}_{Y_t} = \sum_{j=1}^{M} y_{j,t} y'_{j,t}.$$

We assume $d > M$; thus, the realized covariance matrix is estimated under this assumption.

2.3.1 POET Method

The eigenvalues of the realized covariance matrix $\hat{\Sigma}_{Y_t}$ are $\hat{\lambda}_1^t > \hat{\lambda}_2^t > \cdots > \hat{\lambda}_d^t$, and $\hat{\xi}_1^t, \hat{\xi}_2^t, \ldots, \hat{\xi}_d^t$ denote the corresponding eigenvectors. If $\hat{r}$ is the estimator of r, which is the number of factors, $\hat{\Sigma}_{Y_t}$ has a spectral decomposition as follows:

$$\hat{\Sigma}_{Y_t} = \sum_{j=1}^{\hat{r}} \hat{\lambda}_j^t \hat{\xi}_j^t \hat{\xi}_j^{t\prime} + \hat{\Sigma}_{Z_t}, \tag{5}$$

where $\hat{\Sigma}_{Z_t}$ is the covariance matrix of the residual process, which is calculated by $\hat{\Sigma}_{Z_t} = \sum_{j=\hat{r}+1}^{d} \hat{\lambda}_j^t \hat{\xi}_j^t \hat{\xi}_j^{t\prime}$. Here, even if the common factor X_t is an unobservable process, if Σ_Z is sufficiently sparse, $\beta \Sigma_{X_t} \beta'$ in Equation (2) can be estimated using the eigenvalues and eigenvectors of $\hat{\Sigma}_{Y_t}$ [24]. Therefore, we estimate the sparse residual covariance matrix, and then estimate a high-dimensional covariance matrix $\hat{\Sigma}_{Y_t}^S$ as follows:

$$\hat{\Sigma}_{Y_t}^S = \sum_{j=1}^{\hat{r}} \hat{\lambda}_j^t \hat{\xi}_j^t \hat{\xi}_j^{t\prime} + \hat{\Sigma}_{Z_t}^S, \tag{6}$$

where $\hat{\Sigma}_{Z_t}^S$ is the estimated sparse residual covariance matrix. This high-dimensional covariance estimator consists of the POET for low-frequency data of [19] and the PCA approach adopted in [24, 25] for high-frequency data.

2.3.2 Shrinkage POET Method

The PCA which is used in Equation (5) is effective, when dimension d is fixed and the sample size (the number of observations in a day) is sufficiently large. However, it is well-known that in situations where $d > M$, the eigenvalues and eigenvectors of the realized covariance matrix are not consistent estimators in the sense that they are quite far from the true values [16]. To deal with this problem, we use shrinkage POET (SPOET), proposed by [31], which corrects biases of empirical eigenvalues and estimates an HDCM as follows:

$$\tilde{\Sigma}_{Y_t}^S = \sum_{j=1}^{r} \tilde{\lambda}_j^t \hat{\xi}_j^t \hat{\xi}_j^{t\prime} + \hat{\Sigma}_{Z_t}^S,$$

where $\tilde{\lambda}_j^t = \max\{\hat{\lambda}_j^t - \bar{c}d/M, 0\}$. In addition, as $\bar{c}$ is unknown, we have to estimate it. In this paper, we follow [31] to estimate as follows:

$$\hat{c} = (\mathrm{tr}(\hat{\Sigma}_Y^t) - \sum_{j=1}^{r} \hat{\lambda}_j)/(d - r - dr/M). \tag{7}$$

3 Forecasting Models

In this section, in order to forecast an HDCM, we introduce forecasting models based on the PCA. We denote the eigenvalues as:

$$\sigma_f^t = [\lambda_1^t, \ldots, \lambda_r^t]'.$$

Since these eigenvalues are the variances of factors, we can consider models similar to the time-series model of the realized variance of asset returns [29]. To model the eigenvalues, we use the exponentially weighted moving average (EWMA), (Vector) HAR, and (Vector) AR models, the same as [29]. All models except the EWMA model can be easily estimated using OLS.

3.1 EWMA Model

In this paper, we use the EWMA model developed by [33] as a benchmark model, as follows:

$$\sigma_f^{t+1|t} = a\sigma_f^{t|t-1} + (1-a)\sigma^t,$$

where a is the decaying parameter that determines the weight of the observed value 1 period before the forecast, and we set $a = 0.94$ following the framework of a RiskMetrics approach [33]. As this model is easy to implement to forecast volatility and covariance, a lot of studies use it in practice.

3.2 VAR Model

We introduce the AR(1) and VAR models based on high-frequency factor model as:

$$\lambda_i^t = a_{0,i} + a_{1,i}\lambda_i^{t-1} + \varepsilon_i^t, \;\; i = 1, \dots, r, \tag{8}$$

$$\sigma_f^t = A_0 + A_1\sigma_f^{t-1} + \varepsilon_f^t, \tag{9}$$

where $a_{k,i}, A_k, k = 0, 1$ are scalar parameters and parameter matrices, respectively. ε_i^t denotes the innovation term.

Andersen et al. [34] pointed out that the logarithmic standard deviations are closer to a normal distribution in general compared to the realized variance itself, and modeling and forecasting log volatility guarantee that the fitted and forecasted volatility are non-negative without any constrains. Therefore, we also apply the logarithmic eigenvalues to these models.

3.3 V-HAR Model

In this subsection, we introduce the HAR model and V-HAR model which are proposed by [7, 35], respectively. These models are usually applied to forecasting both univariate and multivariate realized volatility. The HAR model is advantaged for approximating the long memory properties using daily, weekly and monthly volatility. Also, given the multivariate framework, the impact of the short- and long-term volatility of another asset can be included in a forecast of the volatility of one asset. To use these models, we calculate the weekly and monthly eigenvalues as follows:

$$\lambda_{i,W}^{t} = \frac{1}{5}\sum_{j=0}^{4}\lambda_{i}^{t-j}, \tag{10}$$

$$\lambda_{i,M}^{t} = \frac{1}{22}\sum_{j=0}^{21}\lambda_{i}^{t-j}. \tag{11}$$

In addition, we define that $\sigma_{f.W}^{t} = [\lambda_{1,W}^{t}, \ldots, \lambda_{r,W}^{t}]'$ and $\sigma_{f.M}^{t} = [\lambda_{1,M}^{t}, \ldots, \lambda_{r,M}^{t}]'$. We construct the HAR and V-HAR models using daily, weekly, and monthly eigenvalues as follows:

$$\lambda_{i}^{t} = a_{0,i} + a_{1,i}\lambda_{i}^{t-1} + a_{2,i}\lambda_{i,W}^{t-1} + a_{3,i}\lambda_{i,M}^{t-1} + \varepsilon_{i}^{t}, \ \ i = 1, \ldots, r, \tag{12}$$

$$\sigma_{f}^{t} = A_0 + A_1\sigma_{f}^{t-1} + A_2\sigma_{f,W}^{t-1} + A_3\sigma_{f,M}^{t-1} + \varepsilon_{f}^{t}. \tag{13}$$

where $a_{k,i}, A_k, k = 0, \ldots, 3$ are scalar parameters and parameter matrices, respectively. Similar to AR and VAR models, these models are transformed into logarithmic models.

Using these models, we can obtain the forecasted HDCM, $\hat{S}_{t+1}$, as follows:

$$\hat{S}_{t+1} = \sum_{j=1}^{\hat{r}} \check{\lambda}_{j}^{t+1}\hat{\xi}_{j}^{t}\hat{\xi}_{j}^{t\prime} + \hat{\Sigma}_{Z_t}^{S}, \tag{14}$$

where $\check{\lambda}_{j}^{t+1}$ denotes the forecasted j-th eigenvalues at $t+1$, $\hat{\xi}_j$ is the eigenvectors corresponding to the forecasted eigenvalues, and $\hat{\Sigma}_{Z_t}^{S}$ is the sparse residual covariance matrix at t. Hence, in this model, we use the eigenvectors and sparse residual covariance matrix at t and the forecasted eigenvalues at $t+1$ to forecast an HDCM.

This study is partly supported by the Institute of Statistical Mathematics (ISM) cooperative research program (2022-ISMCRP-2024), JSPS KAKENHI Grant Number 21K01433, and Grant-in-Aid for JSPS Fellows Grant Number 22J10285. The views expressed in this study are those of the authors and do not necessarily reflect the official views of Nomura Holdings, Inc. or Kwansei Gakuin University.

References

[1] Bauwens, L.; Laurent, S.; Rombouts, J. Multivariate GARCH models: A survey. *J. Appl. Econom.* **2006**, *21*, 79–109.

[2] Engle, R.; Kroner, K. Multivariate simultaneous generalized arch. *Econom. Theory* **1995**, *11*, 122–150.

[3] Tse, Y.; Tsui, A. A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *J. Bus. Econ. Stat.* **2002**, *20*, 351–362.

[4] Engle, R. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *J. Bus. Econ. Stat.* **2002**, *20*, 339–350.

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

[6] Barndorff-Nielsen, O.; Hansen, P.; Lunde, A.; Shephard, N. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econom.* **2011**, *162*, 149–169.

[7] Bubák, V.; Kočenda, E.; Žikeš, F. Volatility transmission in emerging European foreign exchange markets. *J. Bank. Financ.* **2011**, **35**, 2829–2841.

[8] Golosnoy, V.; Gribisch, B.; Liesenfeld, R. The conditional autoregressive Wishart model for multivariate stock market volatility. *J. Econom.* **2012**, *167*, 211–223.

[9] Bauwens, L.; Storti, G.; Violante, F. Dynamic conditional correlation models for realized covariance matrices. *CORE Work. Pap.* **2012**.

[10] Engle, R.; Ledoit, O.; Wolf, M. Large dynamic covariance matrices. *J. Bus. Econ. Stat.* **2019**, *37*, 363–375.

[11] Nakagawa, K.; Imamura, M.; Yoshida, K. Risk-based portfolios with large dynamic covariance matrices. *Int. J. Financ. Stud.* **2018**, *6*, 52.

[12] Moura, G.; Santos, A.; Ruiz, E. Comparing high-dimensional conditional covariance matrices: Implications for portfolio selection. *J. Bank. Financ.* **2020**, *118*, 105882.

[13] De Nard, G.; Engle, R.; Ledoit, O.; Wolf, M. Large dynamic covariance matrices: Enhancements based on intraday data. *J. Bank. Financ.*. **2022**, *138*, 106426.

[14] Trucíos, C.; Mazzeu, J.; Hallin, M.; Hotta, L.; Valls, Pereira, P.; Zevallos, M. Forecasting conditional covariance matrices in high-dimensional time series: A general dynamic factor approach. *J. Bus. Econ. Stat.* **2021**, 1–13.

[15] Wang, Y.; Zou, J. Vast volatility matrix estimation for high-frequency financial data. *Ann. Stat.* **2010**, *38*, 943–978.

[16] Tao, M.; Wang, Y.; Yao, Q.; Zou, J. Large volatility matrix inference via combining low-frequency and high-frequency approaches. *J. Am. Stat. Assoc.* **2011**, *106*, 1025–1040.

[17] Kim, D.; Wang, Y.; Zou, J. Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. *Stoch. Process. Their Appl.* **2016**, *126*, 3527–3577.

[18] Shen, K.; Yao, J.; Li, W. Forecasting high-dimensional realized volatility matrices using a factor model. *Quant. Financ.* **2020**, *20*, 1879–1887.

[19] Fan, J.; Liao, Y.; Mincheva, M. Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2013**, *75*, 603–680.

[20] Fan, J.; Furger, A.; Xiu, D. Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. *J. Bus. Econ. Stat.* **2016**, *34*, 489–503.

[21] Brownlees, C.; Nualart, E.; Sun, Y. Realized networks. *J. Appl. Econom.* **2018**, *33*, 986–1006.

[22] Koike, Y. De-biased graphical lasso for high-frequency data. *Entropy* **2020**, *22*, 1–38.

[23] Fan, J.; Fan, Y.; Lv, J. High dimensional covariance matrix estimation using a factor

model. *J. Econom.* **2008**, *147*, 186–197. http://dx.doi.org/10.1016/j.jeconom.2008.09.017.

[24] Aït-Sahalia, Y.; Xiu, D. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *J. Econom.* **2017**, *201*, 384–399.

[25] Dai, C.; Lu, K.; Xiu, D. Knowing factors or factor loadings, or neither? Evaluating estimators of large covariance matrices with noisy and asynchronous data. *J. Econom.*. **2019**, *208*, 43–79.

[26] Zou, H. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429.

[27] Fan, J.; Li, R. Variable selection via nonconcave penalized. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360.

[28] Cai, T.; Liu, W. Adaptive thresholding for sparse covariance matrix estimation. *J. Am. Stat. Assoc.* **2011**, *106*, 672–684. https://doi.org/10.1198/jasa.2011.tm10560.

[29] Jian, Z.; Deng, P.; Zhu, Z. High-dimensional covariance forecasting based on principal component analysis of high-frequency data. *Econ. Model.* **2018**, *75*, 422–431.

[30] Yata, K.; Aoshima, M. Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *J. Multivar. Anal.* **2012**, *105*, 193–215.

[31] Wang, W.; Fan, J. Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Stat.* **2017**, *45*, 1342–1374.

[32] Rothman, A.; Levina, E.; Zhu, J. Generalized thresholding of large covariance matrices. *J. Am. Stat. Assoc.* **2009**, *104*, 177–186.

[33] Morgan, J.P. *Risk Metrics Thechnical Document*, 4th ed., **1996** (J.P.Morgan/Reuters: New York).

[34] Andersen, T.; Bollerslev, T.; Diebold, F. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Rev The . Econ. Stat.* **2007**, *89*, 701–720.

[35] Corsi, F. A simple approximate long-memory model of realized volatility. *J. Financ. Econom.* **2009**, *7*, 174–196.

[36] Andersen, T.; Dobrev, D.; Schaumburg, E. Jump-robust volatility estimation using nearest neighbor truncation. *J. Econom.* **2012**, *169*, 75–93.

[37] Bollerslev, T.; Patton, A.; Quaedvlieg, R. Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions. *J. Econom.* **2018**, *207*, 71–91.

[38] Diebold, F.; Mariano, R. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–265.

[39] Hansen, P.R.; Lunde, A.; Nason, J.M. The Model Confidence Set. *Econometrica* **2011**, *79*, 453–497.

[40] Laurent, S.; Rombouts, J.; Violante, F. On loss functions and ranking forecasting performances of multivariate volatility models. *J. Econom.* **2013**, *173*, 1–10.